Tikhonov-Phillips Regularizations in Linear Models with Blurred Design

Yu. Golubev* and Th. Zimolo†

Abstract

The paper deals with recovering an unknown vector $\beta \in \mathbb{R}^p$ based on the observations $Y = X\beta + \epsilon\xi$ and $Z = X + \sigma\zeta$, where X is an unknown $n\times p$ - matrix with $n \geq p$, $\xi \in \mathbb{R}^p$ is a standard white Gaussian noise, ζ is a $n \times p$ - matrix with i.i.d. standard Gaussian entries, and $\epsilon, \sigma \in \mathbb{R}^+$ are known noise levels. It is assumed that X has a large condition number and p is large. Therefore, in order to estimate β, the simple Tikhonov-Phillips regularization (ridge regression) with a data-driven regularization parameter is used. For this estimation method, we study the effect of noise $\sigma\zeta$ on the quality of the recovering of $X\beta$ using concentration inequalities for the prediction error.

Keywords: ridge regression, Tikhonov-Phillips method, data-driven regularization, blurred design, prediction mean squared error, concentration inequality.

2000 Mathematics Subject Classification: Primary 62C99; secondary 62C10, 62C20, 62J05.

1 Introduction and main results

This paper deals with recovering an unknown vector $\beta \in \mathbb{R}^p$ based on the noisy observations

$$\begin{aligned} Y &= X\beta + \epsilon\xi, \\ Z &= X + \sigma\zeta, \end{aligned} \tag{1}$$

where X is an unknown $n \times p$ - matrix with $n \geq p$, $\xi \in \mathbb{R}^n$ is a standard white Gaussian noise, ζ is a $n \times p$ - matrix with i.i.d. standard Gaussian entries. For simplicity it is assumed that noise levels $\epsilon,\ \sigma > 0$ are known.

Despite a very simple probabilistic structure of this model, estimating β is a non-trivial statistical problem even in the simple case $p = n = 1$ (see

*CNRS, Aix-Marseille Université, I2M, UMR 7373, 13453 Marseille, France, and Institute for Information Transmission Problems, Moscow, Russia

†Aix-Marseille Université, I2M, UMR 7373, 13453 Marseille, France.

[11]), but real difficulties arise when the condition number of X and p are large. In this case, the maximum likelihood estimate of β has typically a large risk and to avoid this drawback one has to make use of an available a priori information about β and X. Recently some new methods of estimating β in this situation have been proposed. These approaches rely on specific statistical interpretations of (1) ranging from "smooth" models for β [6], [13], [15], [5] to "sparse" ones [12], [1], [2].

In this paper, we focus solely on commonly used in practice estimates based on Tikhonov-Phillips regularization technique [17], often called *ridge regression* in statistics. This method relies on the hypothesis that the Euclidean norm of β is bounded and it is usually used when X is known precisely ($\sigma = 0$). The standard ridge regression estimate of β is computed as follows:

$$\hat{\beta}^{\alpha}(Y,X) = \arg\min_{\beta}\Big\{\|Y - X\beta\|^2 + \alpha\|\beta\|^2\Big\} = (X^{\top}X + \alpha I)^{-1}X^{\top}Y,$$

where here and below $\|\cdot\|$ is the Euclidean norm, I is the identity matrix, and $\alpha \in \mathbb{R}^+$ is the regularization parameter. We emphasize that despite the well-known drawbacks of this regularization technique (see e.g. [7]), it remains widely used in practice because of the very low numerical complexity.

In applications, the regularization parameter is usually data-driven and very often [8] it is computed as a minimizer of the unbiased risk estimate of the mean square prediction risk, i.e.

$$\hat{\alpha}(Y,X) = \arg\min_{\alpha\ge 0}\Big\{\big\|Y - X\hat{\beta}^{\alpha}(Y,X)\big\|^2 + 2\epsilon^2\mathrm{tr}\big[H^{\alpha}(X)\big]\Big\}, \qquad (2)$$

where

$$H^{\alpha}(X) = X(X^{\top}X + \alpha I)^{-1}X^{\top}.$$

Thus, we estimate β with the help of $\hat{\beta}^{\hat{\alpha}(Y,X)}(Y,X)$. From a mathematical viewpoint, the most elegant and interesting fact about performance of this method goes back to Kneip [14].

Theorem 1 *For any $x \ge 1$*

$$\mathbf{P}\Big\{\|X\hat{\beta}^{\hat{\alpha}(Y,X)}(Y,X) - X\beta\| \ge \sqrt{\min_{\alpha\ge 0} R^{\alpha}(X,\beta)} + \epsilon x\Big\} \le \exp(-Kx^2), \qquad (3)$$

where here and in what follows K stands for generic constants and

$$\begin{aligned} R^{\alpha}(\beta,X) &\stackrel{\mathrm{def}}{=} \mathbf{E}\|X\hat{\beta}^{\alpha}(Y,X) - X\beta\|^2 \\ &= \big\|[I - H^{\alpha}(X)]X\beta\big\|^2 + \epsilon^2\mathrm{tr}\big[H^{\alpha}(X)^{\top}H^{\alpha}(X)\big] \end{aligned}$$

is the mean square prediction risk.

In order to understand why this theorem is really surprising, let us chose α as a minimizer of the prediction error, i.e.

$$\alpha(\beta, X) = \arg\min_{\alpha \ge 0} R^{\alpha}(\beta, X). \tag{4}$$

Notice that $\alpha(\beta, X)$ depends on β, so it is not a statistical estimate in the ordinary sense and it cannot be used in practice. It can be interpreted solely as an oracle regularization parameter. One can check with a rather simple algebra that for sufficiently large x

$$\mathbf{P}\Big\{\|X\hat\beta^{\alpha(X,\beta)}(Y, X) - X\beta\| \ge \sqrt{R^{\alpha(\beta,X)}(\beta, X)} + \epsilon x\Big\} \ge \exp(-K' x^2), \tag{5}$$

where K' is a constant. So, comparing Equations (5) and (3), we see that the booth prediction errors $\|X\hat\beta^{\hat\alpha(Y,X)}(Y, X) - X\beta\|$ and $\|X\hat\beta^{\alpha(\beta,X)}(Y, X) - X\beta\|$ are close to $\sqrt{\min_{\alpha \ge 0} R^{\alpha}(\beta, X)}$ at the same parametric rate ϵ. This is why Equation (3) is often called *oracle inequality.*

We emphasize that if the risk of an estimate $\tilde\beta(Y, X)$ is measured by $\|\tilde\beta(Y, X) - \beta\|$, then mathematical analysis of this risk becomes rather complicated. For instance, to our knowledge nothing similar to Theorem 1 is known about the distribution of $\|\hat\beta^{\hat\alpha(Y,X)}(Y, X) - \beta\|$ despite the fact that $\hat\beta^{\hat\alpha(Y,X)}(Y, X)$ behaves reasonably good in numerous applications [8]. The principal difficulties in obtaining good upper bounds for $\|\tilde\beta^{\alpha}(Y, X) - \beta\|$ are due to the data-driven choice of regularization parameters. During last years, several new approaches to this fundamental statistical problem have been proposed (see e.g. [3], [9]), but unfortunately, these methods need the Singular Value Decomposition of X, thus resulting in hardly computable statistical methods compared to (2).

The main goal in this paper is to study the case where the design matrix X is blurred by a white Gaussian noise and to find out how this noise affects the quality of recovering $X\beta$. In order to estimate β in this situation, we make use of the simple plug-in ridge regression estimate

$$\hat\beta^{\alpha}(Y, Z) = \arg\min_{\beta}\Big\{\|Y - Z\beta\|^2 + \alpha\|\beta\|^2\Big\} = (Z^\top Z + \alpha I)^{-1} Z^\top Y \tag{6}$$

combined with the data-driven regularization parameter

$$\hat\alpha(Y, Z) = \arg\min_{\alpha \ge \alpha_\circ}\Big\{\|Y - Z\hat\beta^{\alpha}(Y, Z)\|^2 + 2\epsilon^2 \mathrm{tr}[H^{\alpha}(Z)]\Big\}, \tag{7}$$

where $\alpha_\circ > 0$ is given, and $H^{\alpha}(Z) = Z(Z^\top Z + \alpha I)^{-1} Z^\top$.

In what follows, we are interested in the concentration properties of the empirical analog of the prediction error $\|Z\hat\beta^{\alpha}(Y, Z) - X\beta\|$.

The next theorem shows that when Z is given, this random variable is concentrated in the vicinity of $\sqrt{\min_{\alpha \ge \alpha_\circ} R^{\alpha}(\beta, Z)}$, where

$$\begin{aligned} R^{\alpha}(\beta, Z) &\stackrel{\text{def}}{=} \mathbf{E}_Z\|Z\hat\beta^{\alpha}(Y, Z) - X\beta\|^2 \\ &= \big\|[I - H^{\alpha}(Z)]X\beta\big\|^2 + \epsilon^2 \mathrm{tr}\big[H^{\alpha}(Z)^\top H^{\alpha}(Z)\big]. \end{aligned}$$

Here and in what follows $\mathbf{E}_Z$ and $\mathbf{P}_Z$ stand for the expectation and probability measure generated by the observations $\{Y, Z\}$ from (1) when Z is a given matrix.

Theorem 2 *For any $\lambda \in \mathbb{R}^+$ and any $\alpha \geq \alpha_\circ$*

$$\mathbf{E}_Z \exp\Bigl\{\lambda \bigl\| Z\hat{\beta}^{\hat{\alpha}(Y,Z)}(Y,Z) - X\beta \bigr\|\Bigr\} \leq \exp\Bigl\{\lambda\sqrt{R^\alpha(\beta, Z)} + K\epsilon^2\lambda^2\Bigr\}. \quad (8)$$

This theorem permits to control the concentration of the empirical prediction error of the data-driven ridge regression estimate via the concentration of $R^\alpha(\beta, Z)$ for given α. More precisely, it says that for any $\alpha \geq \alpha_\circ$

$$\|Z\hat{\beta}^{\hat{\alpha}(Y,Z)}(Y,Z) - X\beta\| \overset{\mathbf{P}}{\leq} \sqrt{R^\alpha(\beta, Z)} + K\epsilon\xi_+,$$

where ξ_+ is a positive sub-Gaussian random variable and symbol $\overset{\mathbf{P}}{\leq}$ means that this inequality is interpreted similar to (3).

Therefore, our next step is to study concentration of $R^\alpha(\beta, Z)$. Unfortunately, in spite of the apparent simplicity, this problem is rather difficult from a mathematical viewpoint. Therefore, in this paper, we focus solely on the second order approximation (with respect to σ) of $R^\alpha(\beta, Z)$. In other words, we will focus on the random variable $\widetilde{R}^\alpha_\sigma(\beta, X, \zeta)$ defined by

$$R^\alpha(\beta, X + \sigma\zeta) = \widetilde{R}^\alpha_\sigma(\beta, X, \zeta) + o(\sigma^2), \quad \text{as} \quad \sigma \to 0. \quad (9)$$

In order to make concentration inequalities simpler and more transparent, we will assume the following condition.

Condition 1 *There exists a constant $Q \geq 1$ such that for any $\alpha \in \mathbb{R}^+$*

$$\mathrm{tr}\bigl[H^\alpha(X)\bigr] \leq Q\,\mathrm{tr}\bigl[H^\alpha(X)^\top H^\alpha(X)\bigr].$$

Roughly speaking, this condition means that eigenvalues $\lambda_k(X^\top X)$ of $X^\top X$ decrease rather rapidly with k. We emphasize that only in this case Tikhonov's regularization may have significant advantages with respect to the maximum likelihood estimate.

We begin our study of $\widetilde{R}^\alpha_\sigma(\beta, X, \zeta)$ with the simple case assuming that X in (1) is a diagonal matrix. So, we have at out disposal the noisy observations

$$\begin{aligned} Y_i &= X_i\beta_i + \epsilon\xi_i, \\ Z_i &= X_i + \sigma\zeta_i, \quad i = 1, \ldots, p, \end{aligned} \quad (10)$$

where ζ, ξ are independent standard white Gaussian noises, and we estimate $\beta \in \mathbb{R}^p$ with the help of the ridge regression method.

Notice that this statistical model may be viewed as a simple probabilistic model describing the famous blind deconvolution problem (see e.g. [4, 11]).

Theorem 3 *Suppose X in (1) is a diagonal matrix and Condition 1 holds. Then for any given $\alpha \geq \alpha_\circ$ and any $\lambda \in \mathbb{R}^+$*

$$\begin{aligned}\mathbf{E}\exp\Big\{\lambda\sqrt{[\widetilde{R}_\sigma^\alpha(\beta,X,\zeta)]_+}\Big\} \leq \exp\bigg\{\lambda\bigg[\Big(1+\frac{KQ\sigma^2}{\alpha}\Big)R^\alpha(\beta,X)\bigg]^{1/2} \\ +\frac{9\sigma^2\lambda^2}{8}\max\bigg[\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\bigg]\bigg\}.\end{aligned} \tag{11}$$

Let us briefly discuss this result. If we choose

$$\alpha = \alpha^\circ(\beta,X) = \arg\min_{\alpha\geq\alpha_\circ} R^\alpha(\beta,X),$$

then we obtain from (11) the following inequality:

$$\begin{aligned}\mathbf{E}\exp\Big\{\lambda\sqrt{\big[\widetilde{R}_\sigma^{\alpha^\circ}(\beta,X,\zeta)\big]_+}\Big\} \leq \exp\bigg\{\lambda\bigg[\Big(1+\frac{KQ\sigma^2}{\alpha_\circ}\Big)R^{\alpha^\circ}(\beta,X)\bigg]^{1/2} \\ +\frac{9\sigma^2\lambda^2}{8}\max\bigg[\frac{\epsilon^2}{\alpha_\circ},\|\beta\|_\infty^2\bigg]\bigg\}.\end{aligned} \tag{12}$$

Thus, we see that compared to (8) the noise in X results in

- strengthening of the oracle risk by the factor of $1 + KQ\sigma^2/\alpha_\circ$;
- shifting of the "mean" prediction error by

  $$\frac{3}{2}\max\Big\{\frac{\epsilon}{\sqrt{\alpha_\circ}},\|\beta\|_\infty\Big\}\sigma\xi_+,$$

  where ξ_+ is a positive sub-Gaussian random variable.

Remark 1. $\alpha_\circ$ (see (7)) plays a rather important role when the design is blurred. It follows from (12) that in order to guarantee the global concentration rate similar to (8), $\alpha_\circ$ must be of order of ϵ^2 and σ^2 of order of ϵ^2.

In the general case, the concentration inequality for $\big[\widetilde{R}_\sigma^\alpha(\beta,X,\zeta)\big]_+$ has a little bit more complicated form.

Theorem 4 *Suppose Condition 1 holds. Then for any given $\alpha \geq \alpha_\circ$ and any $\lambda \in \mathbb{R}^+$*

$$\begin{aligned}&\mathbf{E}\exp\Big\{\lambda\sqrt{[\widetilde{R}_\sigma^\alpha(\beta,X,\zeta)]_+}\Big\} \leq \\ &\leq \exp\bigg\{\lambda\bigg[\Big(1+\frac{KQp\sigma^2}{\alpha}\Big)R^\alpha(\beta,X)+\frac{\sigma^2pS^\alpha(\beta,X)}{\alpha^2}\bigg]^{1/2} \\ &+\frac{K\lambda^2\sigma^2}{\alpha}\bigg[R^\alpha(\beta,X)+\epsilon^2+\frac{S^\alpha(\beta,X)}{\alpha}\bigg]\bigg\}.\end{aligned} \tag{13}$$

where $S^\alpha(\beta,X) = \big\|X^\top\big[I-H^\alpha(X)\big]X\beta\big\|^2$.

Remark 2. It is easy to see that $S^\alpha(\beta, X)$ can be bounded from above as follows:

$$S^\alpha(\beta, X) = \alpha^2 \|X^\top(\alpha I + X^\top X)^{-1} X\beta\|^2 \le \alpha^2\|\beta\|^2.$$

Remark 3. Equation (13) reveals the principal drawback of the naive plug-in approach. This method works good solely when $p\sigma^2$ is small, more precisely, when this value is of order of the oracle regularization parameter (4). However, we emphasize here once again that important advantage of this method is related to its low numerical complexity.

In order to improve the low concentration of the plug-in approach, we have to incorporate into the estimate of $(X^\top X)^{-1}$ an a priori information about X. This may be done in different ways. For instance, in [4] it is assumed that

$$X = \sum_{k=1}^{p} \sqrt{b_k}\varphi_k \varphi_k^{*\top},$$

where

- $b_k \in \mathbb{R}^+$ are unknown, but such that $b_1 \ge b_2 \ge \ldots \ge b_p$;
- φ_k and φ_k^* are known orthonormal systems in $\mathbb{R}^n$ and $\mathbb{R}^p$ respectively.

It is easy to check that this assumption reduces (1) to diagonal Model (10). Notice also that this model is equivalent to very specific deconvolution problems (see [11]). For classical statistical linear models, where usually $n \gg p$, the above hypothesis looks rather restrictive.

A less restrictive assumption would be to suppose

$$X^\top X = \sum_{k=1}^{p} b_k \varphi_k \varphi_k^\top, \tag{14}$$

where $b_k \in \mathbb{R}^+$ are unknown and φ_k is a known orthonormal system in R^p.

2 Proofs

2.1 Proof of Theorem 2

The proof of the next simple auxiliary lemma can be found in [10].

Lemma 1 *I) If for some $U, u \ge 0$*

$$U \le u + \min_{\gamma\in(0,F]}\left\{\frac{x^2}{\gamma} + \gamma U\right\}, \quad \textit{then} \quad \sqrt{U} \le \sqrt{r} + |x| \max\left\{2, \sqrt{\frac{2}{F}}\right\}.$$

II) If for some $U, u \ge 0$

$$U \le u + \min_{\gamma\in(0,F]}\left\{\frac{x^2}{\gamma} + \gamma u\right\}, \quad \textit{then} \quad \sqrt{U} \le \sqrt{u} + |x| \max\left\{1, \sqrt{\frac{2}{F}}\right\}.$$

2.1.1 Proof of the theorem

Let $\{s_k(Z) \in \mathbb{R}^+,\ \varphi_k(Z) \in \mathbb{R}^p,\ k = 1, \ldots, p\}$ be eigenvalues and eigenvectors of $Z^\top Z$, i.e.,

$$Z^\top Z\varphi_k(Z) = s_k(Z)\varphi_k(Z).$$

It is easy to check that

$$\varphi_k^*(Z) = \frac{Z\varphi_k(Z)}{\sqrt{s_k(Z)}},\ k = 1, \ldots, p,$$

are the orthonormal vectors in $\mathbb{R}^n$ and

$$Z = \sum_{k=1}^{p} \sqrt{s_k(Z)}\varphi_k^*(Z)\varphi_k^\top(Z) \quad ZZ^\top = \sum_{k=1}^{p} \sqrt{s_k(Z)}\varphi_k^*(Z)\varphi_k^{*\top}(Z).$$

Therefore

$$H^\alpha(Z) = (\alpha I + ZZ^\top)^{-1}ZZ^\top = \sum_{k=1}^{p} h_k^\alpha(Z)\varphi_k^*(Z)\varphi_k^{*\top}(Z),$$

where

$$h_k^\alpha(Z) = \frac{s_k(Z)}{\alpha + s_k(Z)}.$$

Projecting the observations Y onto the linear space spanned by vectors $\{\varphi_k^*(Z),\ k = 1, \ldots, p\}$, we arrive at estimating

$$\bar{\mu}_k(Z) \stackrel{\text{def}}{=} \langle X\beta, \varphi_k^*(Z)\rangle, \quad k = 1, \ldots, p$$

based on the observations

$$\bar{Y}_k(Z) \stackrel{\text{def}}{=} \langle Y, \varphi_k^*(Z)\rangle = \bar{\mu}_k(Z) + \epsilon\xi_k', \quad k = 1, \ldots, p. \tag{15}$$

where ξ_k' are i.i.d. standard Gaussian random variables.

It is also easy to check that

$$\|Z\hat{\beta}^\alpha(Y,Z) - X\beta\|^2 = \sum_{k=1}^{p} \big[\bar{\mu}_k(Z) - h_k^\alpha(Z)\bar{Y}_k(Z)\big]^2,$$
$$\|Z\hat{\beta}^\alpha(Y,Z) - Y\|^2 = \sum_{k=1}^{p} \big[\bar{Y}_k(Z) - h_k^\alpha(Z)\bar{Y}_k(Z)\big]^2,$$

Therefore, in view of (15), we have

$$\begin{aligned}\|Z\hat{\beta}^\alpha(Y,Z) - X\beta\|^2 &= L[\bar{\mu}(Z), h^\alpha(Z)] \\ &+ 2\epsilon\sum_{k=1}^{p}\big[1 - h_k^\alpha(Z)\big]h_k^\alpha(Z)\bar{\mu}_k(Z)\xi_k' + \epsilon^2\sum_{k=1}^{p}\big[h_k^\alpha(Z)\big]^2\big[(\xi_k')^2 - 1\big]\end{aligned} \tag{16}$$

and

$$
\begin{aligned}
\|Z\hat{\beta}^{\alpha}(Y,Z)-Y\|^2 =& L[\bar{\mu}(Z),h^{\alpha}(Z)] - 2\epsilon^2\sum_{k=1}^{p} h_k^{\alpha}(Z) + \epsilon^2\sum_{k=1}^{p}\bigl(\xi_k'\bigr)^2 \\
&+ 2\epsilon\sum_{k=1}^{p}\bigl[1-h_k^{\alpha}(Z)\bigr]^2\bar{\mu}_k(Z)\xi_k' \qquad (17)\\
&+ \epsilon^2\sum_{k=1}^{p}\bigl\{\bigl[h_k^{\alpha}(Z)\bigr]^2-2h_k^{\alpha}(Z)\bigr\}\bigl[(\xi_k')^2-1\bigr],
\end{aligned}
$$

where here and in what follows

$$
L[\bar{\mu}(Z),h^{\alpha}(Z)] = \sum_{k=1}^{p}\bigl[1-h_k^{\alpha}(Z)\bigr]^2\bar{\mu}_k^2(Z) + \epsilon^2\sum_{k=1}^{p}\bigl[h_k^{\alpha}(Z)\bigr]^2.
$$

In view of the definition of $\hat{\alpha}(Y,Z)$ (see (7)), we have for any given $\alpha \ge \alpha_\circ$

$$
\begin{aligned}
\|Z\hat{\beta}^{\hat{\alpha}(YZ)}(Y,Z)-Y\|^2 + 2\epsilon^2\sum_{k=1}^{p} h_k^{\hat{\alpha}(Y,Z)}(Y,Z) \le& \|Z\hat{\beta}^{\alpha}(Y,Z)-Y\|^2 \\
&+ 2\epsilon^2\sum_{k=1}^{p} h_k^{\alpha}(Z)
\end{aligned}
$$

and using (17), we obtain the following equivalent form of this inequality:

$$
\begin{aligned}
&L\bigl[\bar{\mu}(Z),h^{\hat{\alpha}(Y,Z)}(Z)\bigr] + 2\epsilon\sum_{k=1}^{p}\bigl[1-h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2\bar{\mu}_k(Z)\xi_k' \\
&\quad + \epsilon^2\sum_{k=1}^{p}\bigl\{\bigl[h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2-2h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr\}\bigl[(\xi_k')^2-1\bigr] \\
&\le L[\bar{\mu}(Z),h^{\alpha}(Z)] + 2\epsilon\sum_{k=1}^{p}\bigl[1-h_k^{\alpha}(Z)\bigr]^2\mu_k(Z)\xi_k' \\
&\quad + \epsilon^2\sum_{k=1}^{p}\bigl\{\bigl[h_k^{\alpha}(Z)\bigr]^2-2h_k^{\alpha}(Z)\bigr\}\bigl[(\xi_k')^2-1\bigr].
\end{aligned}
$$

Let $\gamma,\gamma' \in (0,3/16)$. Then we can rewrite this inequality as follows:

$$
\begin{aligned}
&L\bigl[\bar{\mu}(Z),h^{\hat{\alpha}(Y,Z)}(Z)\bigr] \le L\bigl[\bar{\mu}(Z),h^{\alpha}(Z)\bigr] \\
&\quad + \rho_1(\gamma) + \gamma L\bigl[\bar{\mu}(Z),h^{\hat{\alpha}(Y,Z)}(Z)\bigr] + \rho_2(\gamma) + \gamma L\bigl[\bar{\mu}(Z),h^{\hat{\alpha}(Y,Z)}(Z)\bigr] \qquad (18)\\
&\quad + \rho_3(\gamma') + \gamma' L\bigl[\bar{\mu}(Z),h^{\alpha}(Z)\bigr] + \rho_4(\gamma') + \gamma' L\bigl[\bar{\mu}(Z),h^{\alpha}(Z)\bigr].
\end{aligned}
$$

where

$$
\rho_1(\gamma) = -2\epsilon\sum_{k=1}^{p}\bigl[1-h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2\bar{\mu}_k(Z)\xi_k' - \gamma\sum_{k=1}^{p}\bigl[1-h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2\bar{\mu}_k^2(Z),
$$

$$
\begin{aligned}
\rho_2(\gamma) =& - \epsilon^2 \sum_{k=1}^{p} \bigl\{ \bigl[h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2 - 2h_k^{\hat{\alpha}(Y,Z)}(Z) \bigr\} \bigl[(\xi_k')^2 - 1\bigr] \\
& - \epsilon^2 \gamma \sum_{k=1}^{p} \bigl[h_k^{\hat{\alpha}(Y,Z)}(Z)\bigr]^2, \\
\rho_3(\gamma') =& 2\epsilon \sum_{k=1}^{p} \bigl[1 - h_k^{\alpha}(Z)\bigr]^2 \bar{\mu}_k(Z)\xi_k' - \gamma' \sum_{k=1}^{p} \bigl[1 - h_k^{\alpha}(Z)\bigr]^2 \bar{\mu}_k^2(Z),
\end{aligned}
$$

and

$$
\rho_4(\gamma') = \epsilon^2 \sum_{k=1}^{p} \bigl\{ \bigl[h_k^{\alpha}(Z)\bigr]^2 - 2h_k^{\alpha}(Z) \bigr\} \bigl[(\xi_k')^2 - 1\bigr] - \gamma' \epsilon^2 \sum_{k=1}^{p} \bigl[h_k^{\alpha}(Z)\bigr]^2.
$$

By Lemma 10 we obtain for any integer $m \geq 1$

$$
\bigl[\mathbf{E}\rho_l^m(\gamma)\bigr]^{1/m} \leq \frac{K\epsilon^2 m}{\gamma}, \quad l = 1, \ldots, 4.
$$

Therefore with (18) we arrive at

$$
\begin{aligned}
\bigl\{\mathbf{E}\bigl[L\bigl[\bar{\mu}(Z), h^{\hat{\alpha}(Y,Z)}(Z)\bigr]\bigr]^m\bigr\}^{1/m} \leq & L\bigl[\bar{\mu}(Z), h^{\alpha}(Z)\bigr] \\
& + \frac{K\epsilon^2 m}{\gamma} + \gamma \bigl\{\mathbf{E}\bigl[L\bigl[\bar{\mu}(Z), h^{\hat{\alpha}(Y,Z)}(Z)\bigr]\bigr]^m\bigr\}^{1/m} \\
& + \frac{K\epsilon^2 m}{\gamma'} + \gamma' L\bigl[\bar{\mu}(Z), h^{\alpha}(Z)\bigr].
\end{aligned}
$$

So, minimizing the right-hand side at this equation in $\gamma, \gamma' \in (0, 3/16)$ and using Lemma 1, we obtain that for any integer m

$$
\biggl\{\mathbf{E}\Bigl[\sqrt{L\bigl[\bar{\mu}(Z), h^{\hat{\alpha}(Y,Z)}(Z)\bigr]}\Bigr]^m\biggr\}^{1/m} \leq \sqrt{L\bigl[\bar{\mu}(Z), h^{\alpha}(Z)\bigr]} + K\epsilon\sqrt{m}. \tag{19}
$$

The similar technique is used in proving (see (16)) the following inequality:

$$
\begin{aligned}
\Bigl\{\mathbf{E}_Z \| Z\hat{\beta}^{\hat{\alpha}(Y,Z)}(Y,Z) - X\beta\|^m\Bigr\}^{1/m} \leq & \biggl\{\mathbf{E}_Z\Bigl[\sqrt{L\bigl[\bar{\mu}(Z), h^{\hat{\alpha}(Y,Z)}(Z)\bigr]}\Bigr]^m\biggr\}^{1/m} \\
& + K\epsilon\sqrt{m}.
\end{aligned}
$$

Finally, combining this equation with (19), we finish the proof. ■

2.2 Proof of Theorem 3

Let us first compute $\widetilde{R}_\sigma^\alpha(\beta, X, \zeta)$. We will make use of the following formula:

$$
R^\alpha(\beta, Z) = L[\beta, H^\alpha(Z)] \stackrel{\text{def}}{=} \sum_{k=1}^{p} \Bigl\{ \bigl[1 - H_k^\alpha(Z)\bigr]^2 X_k^2 \beta_k^2 + \epsilon^2 \bigl[H_k^\alpha(Z)\bigr]^2 \Bigr\},
$$

where

$$H_k^\alpha(Z) = \frac{Z_k^2}{\alpha + Z_k^2}.$$

We obviously have

$$\begin{aligned}
\frac{\partial}{\partial Z_k} L[\beta, H^\alpha(Z)] =& \Big\{ 2\epsilon^2 H_k^\alpha(Z) - 2\big[1 - H_k^\alpha(Z)\big] X_k^2 \beta_k^2 \Big\} \frac{\partial}{\partial Z_k} H^\alpha(Z), \\
\frac{\partial^2}{\partial Z_k^2} L[\beta, H^\alpha(Z)] =& \Big\{ 2\epsilon^2 H_k^\alpha(Z) - 2\big[1 - H_k^\alpha(Z)\big] X_k^2 \beta_k^2 \Big\} \frac{\partial^2}{\partial Z_k^2} H_k^\alpha(Z) \\
& + \big(2X_k^2 + 2\epsilon^2\big) \left[\frac{\partial}{\partial Z_k} H^\alpha(Z) \right]^2,
\end{aligned}$$

and

$$\frac{\partial}{\partial Z_k} H_k^\alpha(Z) = \frac{2\alpha Z_k}{(\alpha + Z_k^2)^2}, \quad \frac{\partial^2}{\partial Z_k^2} H_k^\alpha(Z) = \frac{2\alpha(\alpha - 3Z_k^2)}{(\alpha + Z_k^2)^3}.$$

Therefore with the help of Taylor's expansion we arrive at (see (9))

$$\widetilde{R}_\sigma^\alpha(\beta, X, \zeta) = R^\alpha(\beta, X) + \sigma \Delta_1^\alpha(\beta, X, \zeta) + \frac{\sigma^2}{2} \Delta_2^\alpha(\beta, X, \zeta),$$

where

$$\begin{aligned}
\Delta_1^\alpha(\beta, X, \zeta) =& \sum_{k=1}^p \zeta_k \frac{\partial}{\partial X_k} L[\beta, H^\alpha(X)] \\
=& \, 4 \sum_{k=1}^p \zeta_k \Big\{ \epsilon^2 H_k^\alpha(X) - \big[1 - H_k^\alpha(X)\big] X_k^2 \beta_k^2 \Big\} \frac{\big[1 - H_k(X)\big] X_k}{\alpha + X_k^2},
\end{aligned} \tag{20}$$

and

$$\begin{aligned}
\Delta_2^\alpha(\beta, X, \zeta) =& \sum_{k=1}^p \zeta_k^2 \frac{\partial^2}{\partial X_k^2} L[\beta, H^\alpha(X)] \\
=& \sum_{k=1}^p \zeta_k^2 X_k^2 \beta_k^2 \big[1 - H_k^\alpha(X)\big]^2 \frac{7H_k^\alpha(X) - \big[H_X^\alpha(X)\big]^2 - 2}{\alpha + X_k^2} \\
& + 8\epsilon^2 \sum_{k=1}^p \zeta_k^2 H_k^\alpha(X) \big[1 - H_k^\alpha(X)\big] \frac{9 - 12 H_k^\alpha(X)}{\alpha + X_k^2}.
\end{aligned} \tag{21}$$

By (20) we have

$$
\begin{aligned}
\mathbf{E}\big[\Delta_1^\alpha(\beta,X,\zeta)\big]^2 =&\sum_{k=1}^p\Big\{\epsilon^2 H_k^\alpha(X)-\big[1-H_k^\alpha(X)\big]X_k^2\beta_k^2\Big\}^2\frac{\big[1-H_k^\alpha(X)\big]^2X_k^2}{(\alpha+X_k^2)^2}\\
\le&\epsilon^4\sum_{k=1}^p[H_k^\alpha(X)]^2[1-H_k^\alpha(X)]^2\frac{X_k^2}{(\alpha+X_k^2)^2}\\
&+\sum_{k=1}^p\big[1-H_k^\alpha(X)\big]^4(X_k^2\beta_k^2)^2\frac{X_k^2}{(\alpha+X_k^2)^2}\\
\le&\frac{\epsilon^2}{\alpha}\epsilon^2\sum_{k=1}^p[H_k^\alpha(X)]^2+\max_k[1-H_k^\alpha(X)]^2X_k^2\beta_k^2\frac{X_k^2}{(\alpha+X_k^2)^2}\\
&\times\sum_{k=1}^p[1-H_k^\alpha(X)]^2X_k^2\beta_k^2\le\max\Big\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\Big\}R^\alpha(\beta,X).
\end{aligned}
$$

Since $\Delta_1^\alpha(\beta,X,\zeta)$ is a zero mean Gaussian random variable, we obtain from the above equation

$$
\begin{aligned}
&\mathbf{E}\exp\Big[\lambda\sqrt{R^\alpha(\beta,X)+\sigma\big[\Delta_1^\alpha(\beta,X,\zeta)\big]_+}\Big]\\
&\quad\le\exp\Big[\lambda\sqrt{R^\alpha(\beta,X)}\Big]\ \mathbf{E}\exp\Big\{\frac{\lambda\sigma}{2}\max\Big\{\frac{\epsilon}{\sqrt{\alpha}},\|\beta\|_\infty\Big\}\xi_+\Big\}\\
&\quad\le\exp\Big\{\lambda\sqrt{R^\alpha(\beta,X)}+\frac{\lambda^2\sigma^2}{8}\max\Big\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\Big\}\Big\}.
\end{aligned}
\tag{22}
$$

Next, (see (21))

$$
\big[\Delta_2^\alpha(\beta,X,\zeta)\big]_+\le K\sum_{k=1}^p\Big\{\epsilon^2H_k^\alpha(X)+\big[1-H_k^\alpha(X)\big]^2X_k^2\beta_k^2\Big\}\frac{\zeta_k^2}{\alpha+X_k^2}.
$$

Therefore by Condition 1

$$
\mathbf{E}\big[\Delta_2^\alpha(\beta,X,\zeta)\big]_+\le\frac{KQ}{\alpha}R^\alpha(\beta,X). \tag{23}
$$

Next we make use of the following simple algebraic fact:

$$
\max_{x\ge0}\left\{\frac{\alpha^2x}{(\alpha+x)^3}+\frac{\alpha x}{(\alpha+x)^2}\right\}=\frac{2}{3\sqrt{3}}<\frac{1}{2}.
$$

With this equation we get

$$
\begin{aligned}
&\max_k\Big\{\epsilon^2\frac{H_k^\alpha(X)}{\alpha+X_k^2}+\frac{\big[1-H_k^\alpha(X)\big]^2X_k^2}{\alpha+X_k^2}\beta_k^2\Big\}\\
&\le\max\Big\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\Big\}\max_k\Big\{\frac{\alpha X_k^2}{(\alpha+X_k^2)^2}+\frac{\alpha^2X_k^2}{(\alpha+X_k^2)^3}\Big\}\\
&\le\max\Big\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\Big\}\max_{x\ge0}\Big\{\frac{\alpha x}{(\alpha+x)^2}+\frac{\alpha^2x}{(\alpha+x)^3}\Big\}\le\frac{1}{2}\max\Big\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\Big\}
\end{aligned}
$$

and we obtain by (23) and Lemma 11

$$\mathbf{E}\exp\Big[\lambda\sqrt{\sigma^2\big[\Delta_2^\alpha(\beta,X,\zeta)\big]_+}\Big] \le \exp\bigg\{\lambda\sqrt{\frac{KQ\sigma^2 R^\alpha(\beta,X)}{\alpha}} + \frac{\sigma^2\lambda^2}{2}\max\bigg\{\frac{\epsilon^2}{\alpha},\|\beta\|_\infty^2\bigg\}\bigg\}.$$

Finally, combining this equation and (22) with the help of Lemma 12, we complete the proof. ■

2.3 Proof of Theorem 4

First, notice (see (6)) that $\hat{\beta}^\alpha(Y,Z)$ is a solution to

$$Z^\top\big[Y - Z\hat{\beta}^\alpha(Y,Z)\big] = \alpha\hat{\beta}^\alpha(Y,Z). \tag{24}$$

Since we are interested in $Z\hat{\beta}^\alpha(Y,Z)$, we denote for brevity

$$\hat{\mu}^\alpha(Y,Z) = Z\hat{\beta}^\alpha(Y,Z)$$

and we obtain from (24)

$$ZZ^\top\big[Y - \hat{\mu}^\alpha(Y,Z)\big] = \alpha\hat{\mu}^\alpha(Y,Z),$$

or, equivalently,

$$\hat{\mu}^\alpha(Y,Z) = H^\alpha(Z)Y, \quad \text{where} \quad H^\alpha(Z) = (\alpha I + ZZ^\top)^{-1}ZZ^\top.$$

Thus, our problem is reduced to estimating $\mu = X\beta \in \mathbb{R}^n$ based on the noisy data

$$\begin{aligned} Y &= \mu + \epsilon\xi, \\ Z &= X + \sigma\zeta \end{aligned}$$

with the help of the plug-in ridge regression estimates

$$\hat{\mu}^\alpha(Y,Z) = H^\alpha(Z)Y, \quad \alpha \ge \alpha_\circ.$$

In order to compute $\widetilde{R}_\sigma^\alpha(\beta,X,\zeta)$, we will need the second order approximation for $H^\alpha(Z)$ which can be obtained with the help of the following simple formulas:

$$\begin{aligned} H^\alpha(Z) - H^\alpha(X) &= \alpha(\alpha I + XX^\top)^{-1} - \alpha(\alpha I + ZZ^\top)^{-1} \\ &= (\alpha I + XX^\top)^{-1}\big(ZZ^\top - XX^\top\big)[I - H^\alpha(Z)] \\ &= \alpha^{-1}[I - H^\alpha(X)]\big(ZZ^\top - XX^\top\big)[I - H^\alpha(Z)]. \end{aligned}$$

and similarly

$$\begin{aligned}H^\alpha(Z) - H^\alpha(X) &= \alpha(\alpha I + XX^\top)^{-1} - \alpha(\alpha I + ZZ^\top)^{-1}\\&= (\alpha I + XX^\top)^{-1}\big(ZZ^\top - XX^\top\big)[I - H^\alpha(Z)]\\&= \alpha^{-1}[I - H^\alpha(Z)]\big(ZZ^\top - XX^\top\big)[I - H^\alpha(X)].\end{aligned}$$

From the above equations we obtain

$$\begin{aligned}H^\alpha(Z) =&H^\alpha(X) + \alpha^{-1}[I - H^\alpha(X)]\big(ZZ^\top - XX^\top\big)[I - H^\alpha(Z)]\\=&H^\alpha(X) + \alpha^{-1}[I - H^\alpha(X)]\big(ZZ^\top - XX^\top\big)[I - H^\alpha(X)]\\&- \alpha^{-1}[I - H^\alpha(X)]\big(ZZ^\top - XX^\top\big)[H^\alpha(Z) - H^\alpha(X)]\\=&G^\alpha(X,\zeta) + \frac{\sigma^2}{\alpha}[I - H^\alpha(X)]\zeta\zeta^\top[I - H^\alpha(X)]\\&- \alpha^{-2}[I - H^\alpha(X)]\big(ZZ^\top - XX^\top\big)[I - H^\alpha(Z)]\\&\times \big(ZZ^\top - XX^\top\big)[I - H^\alpha(X)],\end{aligned}$$

where

$$G^\alpha(X,\zeta) \stackrel{\text{def}}{=} H^\alpha(X) + \frac{\sigma}{\alpha}[I - H^\alpha(X)]\zeta(X)[I - H^\alpha(X)] \tag{25}$$

is the first order approximation of $H^\alpha(Z)$ and

$$\zeta(X) = X\zeta^\top + \zeta X^\top.$$

So, we arrive at the following second order approximation of $H^\alpha(Z)$:

$$\begin{aligned}\widetilde{H}^\alpha(X,\zeta) =&G^\alpha(X,\zeta) + \frac{\sigma^2}{\alpha}[I - H^\alpha(X)]\zeta\zeta^\top[I - H^\alpha(X)]\\&- \frac{\sigma^2}{\alpha^2}[I - H^\alpha(X)]\zeta(X)[I - H^\alpha(X)]\zeta(X)^\top[I - H^\alpha(X)].\end{aligned} \tag{26}$$

In order to simplify calculations, we will make use of the SVD of X. Let $\lambda_k(X) \in \mathbb{R}^+$, $e_k(X) \in \mathbb{R}^p$, $k = 1, \dots, p$ be eigenvalues and eigenvectors of $X^\top X$, i.e.,

$$X^\top X e_k(X) = \lambda_k(X) e_k(X), \quad k = 1, \dots, p.$$

Define the orthonormal vectors in $\mathbb{R}^n$

$$e_k^*(X) = \frac{Xe_k(X)}{\sqrt{\lambda_k(X)}}, \quad k = 1, \dots, p.$$

It is well-know and easy to check that

$$X = \sum_{k=1}^p \sqrt{\lambda_k(X)} e_k^*(X) e_k^\top(X),$$

and so,

$$
\begin{aligned}
XX^\top &= \sum_{k=1}^{p} \lambda_k e_k^*(X) e_k^{*\top}(X), \\
H^\alpha(X) &= \sum_{k=1}^{p} \frac{\lambda_k(X)}{\alpha + \lambda_k(X)} e_k^*(X) e_k^{*\top}(X).
\end{aligned} \tag{27}
$$

Therefore, denoting,

$$
h_k^\alpha(X) = \frac{\lambda_k(X)}{\alpha + \lambda_k(X)}, \tag{28}
$$

we have

$$
\begin{aligned}
H^\alpha(X) &= \sum_{k=1}^{p} h_k^\alpha(X) e_k^*(X) e_k^{*\top}(X), \\
I - H^\alpha(X) &= \sum_{k=1}^{p} [1 - h_k^\alpha(X)] e_k^*(X) e_k^{*\top}(X).
\end{aligned}
$$

We are ready to obtain the spectral representation of $G^\alpha(X,\zeta)$. Let

$$
\bar{G}_{ij}^\alpha(X,\zeta) = e_i^{*\top}(X) G^\alpha(X,\zeta) e_j^*(X).
$$

Then we get from (25)

$$
\bar{G}_{ij}^\alpha(X,\zeta) = h_i^\alpha(X)\delta_{ij} + \frac{\sigma}{\alpha}\big[1 - h_i^\alpha(X)\big]\big[1 - h_j^\alpha(X)\big]\Xi_{ij}(X),
$$

where

$$
\Xi_{ij}(X) = e_i^{*\top}(X)\zeta(X)e_j^*(X).
$$

The covariance function of this Gaussian random matrix is computed in the following lemma.

Lemma 2

$$
\mathbf{E}\Xi_{ks}(X)\Xi_{k's'}(X) = [\lambda_k(X) + \lambda_s(X)](\delta_{ss'}\delta_{kk'} + \delta_{sk'}\delta_{ks'}). \tag{29}
$$

Proof. Let ζ be a random matrix with i.i.d. entries and B be a non-random matrix. Then

$$
\mathbf{E}\zeta B\zeta = B^\top
$$

since

$$
\sum_{k,l} B_{kl}\mathbf{E}\zeta_{ik}\zeta_{lj} = \sum_{k,l} B_{kl}\delta_{il}\delta_{kj} = B_{ji}.
$$

Similarly

$$
\mathbf{E}\zeta^\top B\zeta^\top = B^\top, \quad \mathbf{E}\zeta B\zeta^\top = I\,\mathrm{tr}(B), \quad \mathbf{E}\zeta^\top B\zeta = I\,\mathrm{tr}(B).
$$

Next, noticing that

$$X = \sum_{l=1}^{p} \sqrt{\lambda_l(X)} e_l^*(X) e_l^\top(X),$$

we get

$$\begin{aligned}\Xi_{ij}(X) =& e_i^{*\top}(X)[X\zeta^\top + \zeta X^\top] e_j^*(X) = \sqrt{\lambda_i(X)} e_i^\top(X)\zeta^\top e_j^*(X) \\ &+ \sqrt{\lambda_j(X)} e_i^{*\top}(X)\zeta e_j(X).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbf{E}\Xi_{ij}(X)\Xi_{kl}(X) =& \mathbf{E}\big[\sqrt{\lambda_i(X)} e_i^\top(X)\zeta^\top e_j^*(X) + \sqrt{\lambda_j(X)} e_i^{*\top}(X)\zeta e_j(X)\big] \\ &\times \big[\sqrt{\lambda_k(X)} e_k^\top(X)\zeta^\top e_l^*(X) + \sqrt{\Lambda_l(X)} e_k^{*\top}(X)\zeta e_l(X)\big] \\ =& \sqrt{\lambda_i(X)\Lambda_k(X)} e_i^\top(X)\mathbf{E}\big[\zeta^\top e_j^*(X) e_k^\top(X)\zeta^\top\big] e_l^*(X) \\ &+ \sqrt{\lambda_i(X)\lambda_l(X)} e_i^\top(X)\mathbf{E}\big(\zeta^\top e_j^*(X) e_k^{*\top}(X)\zeta\big) e_l(X) \\ &+ \sqrt{\lambda_j(X)\lambda_k(X)} e_i^{*\top}(X)\mathbf{E}\big(\zeta e_j(X) e_k^\top(X)\zeta^\top\big) e_l^*(X) \\ &+ \sqrt{\lambda_j(X)\lambda_l(X)} e_i^{*\top}(X)\mathbf{E}\big[\zeta e_j(X) e_k^{*\top}(X)\zeta\big] e_l(X) \\ =& \sqrt{\lambda_i(X)\lambda_k(X)}\delta_{ik}\delta_{jl} + \sqrt{\lambda_i(X)\lambda_l(X)}\delta_{jk}\delta_{il} \\ &+ \sqrt{\lambda_j(X)\lambda_k(X)}\delta_{il}\delta_{jk} + \sqrt{\lambda_j(X)\lambda_l(X)}\delta_{ik}\delta_{jl}. \qquad \blacksquare\end{aligned}$$

In what follows we will need two concentration inequalities related to $\Xi(X)$.

Lemma 3 *Let D be a $p \times p$-diagonal matrix and $u \in \mathbb{R}^p$. Then for any $\lambda \geq 0$*

$$\begin{aligned}\mathbf{E}\exp\Big\{\lambda\big\|D\Xi(X)u\big\|\Big\} \leq& \exp\Big\{\lambda\sqrt{\mathbf{E}\big\|D\Xi(X)u\big\|^2} \\ &+ \lambda^2 \max_j \Big[D_{jj}^2\lambda_j(X)\|u\|^2 + D_{jj}^2\big\|\sqrt{\lambda(X)}u\big\|^2\Big] \\ &+ 2\lambda^2\|D\lambda(X)u\| \times \|Du\|\Big\}.\end{aligned} \tag{30}$$

Proof. For the Gaussian random vector v with components

$$v_i = D_{ii}\sum_{k=1}^{p} \Xi_{ik}(X) u_k$$

one easily obtains (see (29))

$$\begin{aligned}
R_{ij} = \mathbf{E} v_i v_j =& D_{ii} D_{jj} \sum_{k,l=1}^{p} \mathbf{E} \Xi_{ik}(X) \Xi_{jl}(X) u_k u_l \\
=& D_{ii} D_{jj} \sum_{k,l=1}^{p} [\lambda_i(X) + \lambda_k(X)](\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}) u_k u_l \\
=& D_{ii}^2 \delta_{ij} \sum_{k=1}^{p} [\lambda_i(X) + \lambda_k(X)] u_k^2 + D_{ii} D_{jj} [\lambda_i(X) + \lambda_j(X)] u_i u_j.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbf{E}\|v\|^2 = \sum_{k=1}^{p} R_{kk} =& \sum_{k=1}^{p} D_{kk}^2 \lambda_k(X) \sum_{s=1}^{p} u_s^2 + \sum_{k=1}^{p} D_{kk}^2 \sum_{s=1}^{p} u_s^2 \lambda_s(X) \\
&+ 2 \sum_{k=1}^{p} D_{kk}^2 \lambda_k(X) u_k^2.
\end{aligned}$$

Next, we make use of the formula

$$\|v\|^2 = \sum_{k=1}^{p} s_k(R) \xi_k^2,$$

where ξ_k are i.i.d. standard Gaussian random variables and $s_k(u)$ are eigenvalues of R, i.e.

$$R\phi_k = s_k(R)\phi_k, \quad k = 1, \ldots, p.$$

Therefore by Lemma 11

$$\mathbf{E} \exp\bigl\{\lambda\|v\|\bigr\} \leq \exp\Bigl\{\lambda\sqrt{\mathbf{E}\|v\|^2} + \lambda^2 s_1(R)\Bigr\}. \tag{31}$$

In order to bound from above $s_1(R)$, notice that for any $\phi \in \mathbb{R}^p$

$$\begin{aligned}
[R\phi]_i =& D_{ii}^2 \lambda_i(X) \|u\|^2 \phi_i + D_{ii}^2 \phi_i \sum_{k=1}^{p} \lambda_k(X) u_k^2 \\
&+ D_{ii} \lambda_i(X) u_i \sum_{k=1}^{p} D_{kk} u_k \phi_k + D_{ii} u_i \sum_{k=1}^{p} D_{kk} \lambda_k(X) u_k \phi_k.
\end{aligned}$$

Hence, by the Cauchy-Shwarz inequality we obtain from the above equation

$$\begin{aligned}
s_1(R) = \max_{\|\phi\| \leq 1} \|R\phi\| \leq& \max_j \biggl\{ D_{jj}^2 \lambda_j \sum_{k=1}^{p} u_k^2 + D_{jj}^2 \sum_{k=1}^{p} \lambda_k(X) u_k^2 \biggr\} \\
&+ 2 \biggl[\sum_{k=1}^{p} D_{kk}^2 u_k^2\biggr]^{1/2} \biggl[\sum_{k=1}^{p} D_{kk}^2 \lambda_k^2(X) u_k^2\biggr]^{1/2}
\end{aligned}$$

and, so, (30) follows from this inequality and (31). ■

Lemma 4 *Let D be a $p\times p$-diagonal matrix. Then for any $\lambda\in\mathbb{R}^+$*

$$E\exp\Big\{\lambda\sqrt{\operatorname{tr}\big[D\Xi(X)D^2\Xi(X)D\big]}\Big\}\le\exp\Big\{\lambda\sqrt{\mathbf{E}\operatorname{tr}\big[D\Xi(X)D^2\Xi(X)D\big]} + K\lambda^2\max_{i,j}D_{ii}^2D_{jj}^2\lambda_j(X)\Big\}. \tag{32}$$

Proof. Notice

$$\begin{aligned}\operatorname{tr}\big[D\Xi(X)D^2\Xi(X)D\big] &= \sum_{i,j=1}^p D_{ii}^2D_{jj}^2\Xi_{ij}^2(X).\\ &=\sum_{i=1}^p D_{ii}^4\Xi_{ii}^2(X)+2\sum_{j<i}D_{ii}^2D_{jj}^2\Xi_{ij}^2(X).\end{aligned}\tag{33}$$

It is easy to check with the help of (29) that

$$\mathbf{E}\Xi_{ii}(X)\Xi_{jj}(X)=4\lambda_i(X)\delta_{ij},\quad \mathbf{E}\Xi_{ij}(X)\Xi_{ij}(X)=\lambda_i(X)+\lambda_j(X),\ i\neq j,$$

and when $j<i$ and $k<l$

$$\mathbf{E}\Xi_{ij}(X)\Xi_{kl}(X)=[\lambda_i(X)+\lambda_j(X)]\delta_{il}\delta_{kj}=0.$$

So, let $\xi_{ij},\ i,j=1\dots,p$ be i.i.d. standard Gaussian random variables. Then we have

$$\sum_{i=1}^p D_{ii}^4\Xi_{ii}^2(X)\overset{\mathbf{P}}{=}4\sum_{i=1}^p D_{ii}^4\lambda_i(X)\xi_{ii}^2 \tag{34}$$

and

$$\begin{aligned}\sum_{j<i}D_{ii}^2D_{jj}^2\Xi_{ij}^2(X) &=\sum_{j<i}D_{ii}^2D_{jj}^2[\lambda_i(X)+\lambda_j(X)]\xi_{ij}^2\\ &=\sum_{j<i}D_{ii}^2D_{jj}^2\lambda_i(X)\xi_{ij}^2+\sum_{j<i}D_{ii}^2D_{jj}^2\lambda_j(X)\xi_{ij}^2.\end{aligned}\tag{35}$$

Hence by Lemma 11 we obtain

$$\mathbf{E}\exp\bigg\{\lambda\bigg[\sum_{i=1}^p D_{ii}^4\lambda_i(X)\xi_{ii}^2\bigg]^{1/2}\bigg\}\le\exp\bigg\{\lambda\bigg[\sum_{i=1}^p D_{ii}^4\lambda_i(X)\bigg]^{1/2} + \lambda^2\max_i D_{ii}^4\lambda_i(X)\bigg\},$$

$$\mathbf{E}\exp\bigg\{\lambda\bigg[\sum_{j<i}D_{ii}^2D_{jj}^2\lambda_i(X)\xi_{ij}^2\bigg]^{1/2}\bigg\}\le\exp\bigg\{\lambda\bigg[\sum_{j<i}D_{ii}^2D_{jj}^2\lambda_i(X)\bigg]^{1/2} + \lambda^2\max_{i,j}D_{jj}^2D_{ii}^2\lambda_i(X)\bigg\},$$

$$\mathbf{E}\exp\biggl\{\lambda\biggl[\sum_{j<i} D_{jj}^2 D_{ii}^2\lambda_j\xi_{ij}^2\biggr]^{1/2}\biggr\} \le \exp\biggl\{\lambda\biggl[\sum_{j<i} D_{jj}^2 D_{ii}^2\lambda_j(X)\biggr]^{1/2} + \lambda^2\max_{i,j} D_{ii}^2 D_{jj}^2\lambda_j(X)\biggr\}.$$

Therefore, combining the above inequalities and (33)-(35) with the help of Lemma 12, we arrive at (32). ■

In the proof of Theorem 4 the following simple lemma plays an important role.

Lemma 5 *Suppose Condition 1 holds. Then for any given* $\alpha \in \mathbb{R}^+$ *and any* $\mu \in \mathbb{R}^n$

$$\begin{aligned}\mathbf{E}L\bigl[\mu, G^\alpha(X,\zeta)\bigr] \le& \biggl(1+\frac{KQp\sigma^2}{\alpha}\biggr)L\bigl[\mu, H^\alpha(X)\bigr] \\ &+\frac{\sigma^2 p}{\alpha^2}\bigl\|X^\top\bigl[I-H^\alpha(X)\bigr]\mu\bigr\|^2.\end{aligned} \tag{36}$$

where

$$L[\mu, H] = \|(I-H)\mu\|^2 + \epsilon^2\mathrm{tr}\bigl(HH^\top\bigr).$$

Proof. Since $L[\mu, H]$ is a quadratic functional in H, we have (see (25))

$$\begin{aligned}\mathbf{E}L\bigl[\mu, G^\alpha(X,\zeta)\bigr] =& L[\mu, H^\alpha(X)] \\ &+\frac{\sigma^2}{\alpha^2}\mathbf{E}\,\bigl\|[I-H^\alpha(X)]\zeta(X)[I-H^\alpha(X)]\mu\bigr\|^2 \\ &+\frac{\sigma^2\epsilon^2}{\alpha^2}\mathbf{E}\,\mathrm{tr}\Bigl\{\bigl[I-H^\alpha(X)\bigr]\zeta(X)\bigl[I-H^\alpha(X)\bigr] \\ &\times\bigl[I-H^\alpha(X)\bigr]^\top\zeta(X)^\top\bigl[I-H^\alpha(X)\bigr]^\top\Bigr\}.\end{aligned} \tag{37}$$

Let $\bar\mu(X)$ be vector in $\mathbb{R}^p$ with componets $\bar\mu_k(X) = \langle\mu,\, e_k^*(X)\rangle$ and $h(X)$ be the diagonal matrix with the entries $h_k^\alpha(X),\ k=1,\ldots,p$ defined by (28). Then, with the help of Lemma 2, we obtain

$$\begin{aligned}&\mathbf{E}\,\bigl\|[I-H^\alpha(X)]\zeta(X)\bigl[I-H^\alpha(X)\bigr]\mu\bigr\|^2 \\ &\quad=\bar\mu^\top(X)\bigl[I-h^\alpha(X)\bigr]\mathbf{E}\bigl\{\Xi^\top\bigl[I-h^\alpha(X)\bigr]^2\Xi\bigr\}\bigl[I-h^\alpha(X)\bigr]\bar\mu(X) \\ &\quad=\sum_{i=1}^p\bigl[1-h_i^\alpha(X)\bigr]^2\bar\mu_i^2(X)\biggl\{2\bigl[1-h_i^\alpha(X)\bigr]^2\lambda_i(X) \\ &\qquad+\lambda_i(X)\sum_{k=1}^p[1-h_k^\alpha(X)]^2+\sum_{k=1}^p\bigl[1-h_k^\alpha(X)\bigr]^2\lambda_k(X)\biggr\} \\ &\quad\le\alpha(2+p)\sum_{i=1}^p\bigl[1-h_i^\alpha(X)\bigr]^2\bar\mu_i^2(X)+p\sum_{i=1}^p\bigl[1-h_i^\alpha(X)\bigr]^2\bar\mu_i^2(X)\lambda_i(X).\end{aligned}$$

With the similar arguments and Condition 1 we arrive at

$$
\begin{aligned}
&\mathbf{E}\,\mathrm{tr}\Big\{[I-H^\alpha(X)]\zeta(X)[I-H^\alpha(X)][I-H^\alpha(X)]^\top\zeta(X)^\top[I-H^\alpha(X)]^\top\Big\}\\
&\quad=\mathrm{tr}\Big\{[I-h^\alpha(X)]\mathbf{E}\{\Xi[1-h^\alpha(X)]^2\Xi^\top\}[I-h^\alpha(X)]\Big\}\\
&\quad=\sum_{i=1}^p[1-h_i^\alpha(X)]^2\bigg[2[1-h_i^\alpha(X)]^2\lambda_i(X)\\
&\qquad+\lambda_i(X)\sum_{k=1}^p[1-h_k^\alpha(X)]^2+\sum_{k=1}^p[1-h_k^\alpha(X)]^2\lambda_k(X)\bigg]\\
&\quad\le 4Q\alpha p\sum_{i=1}^p[h_i^\alpha(X)]^2.
\end{aligned}
$$

Thus (36) follows from (37) and the above equations. ■

The following lemma controls the concentration of $L[\mu,G^\alpha(X,\zeta)]$.

Lemma 6 *For any given $\alpha>0$ and $\lambda\ge 0$*

$$
\begin{aligned}
\mathbf{E}\exp\Big\{\lambda\sqrt{L\big[\mu,G^\alpha(X,\zeta)\big]}\Big\}&\le\exp\bigg\{\lambda\bigg[\Big(1+\frac{KQp\sigma^2}{\alpha}\Big)L\big[\mu,H^\alpha(X)\big]\\
&\quad+\frac{\sigma^2p}{\alpha^2}\big\|X^\top\big[I-H^\alpha(X)\big]\mu\big\|^2\bigg]^{1/2}\\
&\quad+\frac{K\lambda^2\sigma^2}{\alpha}\Big\{L\big[\mu,H^\alpha(X)\big]+\epsilon^2+\alpha^{-1}\big\|X^\top\big[I-H^\alpha(X)\big]\mu\big\|^2\Big\}\bigg\}.
\end{aligned}
\tag{38}
$$

Proof. Notice that $L\big[\mu,G^\alpha(X,\zeta)\big]$ admits the following decomposition:

$$
L\big[\mu,G^\alpha(X,\zeta)\big]=L\big[\mu,H^\alpha(X)\big]+\frac{2\sigma}{\alpha}\Delta_1^\alpha(\mu,X,\Xi)+\frac{\sigma^2}{\alpha^2}\Delta_2^\alpha(\mu,X,\Xi),\tag{39}
$$

where

$$
\begin{aligned}
\Delta_1^\alpha(\mu,X,\Xi)=&\bar\mu^\top[I-h^\alpha(X)]^2\Xi[I-h^\alpha(X)]^2\bar\mu\\
&+\epsilon^2\mathrm{tr}\Big\{h^\alpha(X)[I-h^\alpha(X)]\Xi[I-h^\alpha(X)]\Big\}
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta_2^\alpha(\mu,X,\Xi)=&\;\bar\mu^\top[I-h^\alpha(X)]\Xi[I-h^\alpha(X)]^2\Xi^\top[I-h^\alpha(X)]\bar\mu\\
&+\epsilon^2\mathrm{tr}\Big\{[I-h^\alpha(X)]\Xi[I-h^\alpha(X)]^2\Xi^\top[I-h^\alpha(X)]\Big\}.
\end{aligned}
$$

Recall that $\bar\mu\in\mathbb{R}^p$ in the above equations has components $\bar\mu_k=\langle\mu,\varphi_k(X)\rangle$.

Notice that $\Delta_1^\alpha[\bar\mu, X, \Xi]$ is a Gaussian random variable with zero mean. Therefore, we need solely to bound from above its variance. We have by Lemma 2

$$
\begin{aligned}
&\mathbf{E}\Big\{\bar\mu^\top[I-h^\alpha(X)]^2\Xi[I-h^\alpha(X)]^2\bar\mu\Big\}^2\\
&\quad=\bar\mu^\top[I-h^\alpha(X)]^2\mathbf{E}\Big\{\Xi[I-h^\alpha(X)]^2\bar\mu\\
&\qquad\times\bar\mu^\top[I-h^\alpha(X)]^2\Xi\Big\}[I-h^\alpha(X)]^2]\bar\mu\\
&\quad=4\sum_{k=1}^p[1-h_k^\alpha(X)]^4\bar\mu_k^2\times\sum_{k=1}^p[1-h_k^\alpha(X)]^4\lambda_k(X)\bar\mu_k^2\\
&\quad\le 4\alpha\bigg[\sum_{k=1}^p[1-h_k^\alpha(X)]^2\bar\mu_k^2\bigg]^2.
\end{aligned}
$$

With the similar arguments we obtain

$$
\begin{aligned}
&\mathbf{E}\Big\{\mathrm{tr}\Big\{h^\alpha(X)[I-h^\alpha(X)]\Xi[I-h^\alpha(X)]\Big\}\Big\}^2\\
&\quad=\mathbf{E}\bigg[\sum_{k=1}^p h_k^\alpha(X)[1-h_k^\alpha(X)]^2\Xi_{kk}\bigg]^2\\
&\quad=\sum_{k=1}^p\sum_{s=1}^p h_k^\alpha(X)[1-h_k^\alpha(X)]^2h_s^\alpha(X)[1-h_s^\alpha(X)]^2\mathbf{E}\Xi_{ss}\Xi_{kk}\\
&\quad=4\sum_{k=1}^p\sum_{s=1}^p h_k^\alpha(X)[1-h_k^\alpha(X)]^2h_s^\alpha(X)[1-h_s^\alpha(X)]^2\lambda_s(X)\delta_{sk}\\
&\quad\le 4\alpha\sum_{k=1}^p[h_k^\alpha(X)]^2.
\end{aligned}
$$

Hence $\mathbf{E}\big[\Delta_1^\alpha(\mu, X, \Xi)\big]^2 \le K\alpha L\big[\mu, H^\alpha(X)\big]$, and thus

$$
\begin{aligned}
&\mathbf{E}\exp\bigg\{\lambda\sqrt{L\big[\mu,H^\alpha(X)\big]+\frac{2\sigma}{\alpha}\big[\Delta_1^\alpha(\mu,X,\Xi)\big]_+}\bigg\}\\
&\qquad\le\exp\bigg\{\lambda\sqrt{L\big[\mu,H^\alpha(X)\big]}+\frac{K\sigma^2\lambda^2}{\alpha}L\big[\mu,H^\alpha(X)\big]\bigg\}.
\end{aligned}
\tag{40}
$$

In order to control $\Delta_2^\alpha(\bar\mu, X, \Xi)$, we make use of Lemmas 3 and 4. With

$$
D=[I-h^\alpha(X)]\quad\text{and}\quad u=[I-h^\alpha(X)]\bar\mu,
$$

we have

$$
\begin{aligned}
&\max_j\Big[D_{jj}^2\lambda_j(X)\|u\|^2+D_{jj}^2\|\sqrt{\lambda(X)}u\|^2\Big]+2\|D\lambda(X)u\|\times\|Du\|\\
&\qquad\le 3\alpha\|[I-h^\alpha(X)]\bar\mu\|^2+\|[I-h^\alpha(X)]\sqrt{\lambda(X)}\bar\mu\|^2,
\end{aligned}
$$

and thus by (30)

$$\begin{aligned}
&\mathbf{E}\exp\Big\{\lambda\big\|[I-h^\alpha(X)]\Xi(X)[I-h^\alpha(X)]\bar\mu\big\|\Big\}\\
&\qquad\le \exp\Big\{\lambda\sqrt{\mathbf{E}\big\|[I-h^\alpha(X)]\Xi(X)[I-h^\alpha(X)]\bar\mu\big\|^2}\\
&\qquad\quad+\lambda^2\big\{3\alpha\|[I-h^\alpha(X)]\bar\mu\|^2+\big\|[I-h^\alpha(X)]\sqrt{\lambda(X)}\bar\mu\big\|^2\big\}\Big\}.
\end{aligned}\tag{41}$$

Similarly, with the help of (32) we obtain

$$\begin{aligned}
&\mathbf{E}\exp\Big\{\lambda\sqrt{\mathrm{tr}\big\{[I-h^\alpha(X)]\Xi(X)[I-h^\alpha(X)]^2\Xi^\top(X)[I-h^\alpha(X)]\big\}}\Big\}\\
&\qquad\le \mathbf{E}\exp\Big\{\lambda\sqrt{\mathbf{E}\mathrm{tr}\big\{[I-h^\alpha(X)]\Xi(X)[I-h^\alpha(X)]^2\Xi^\top[I-h^\alpha(X)]\big\}}\\
&\qquad\quad+K\alpha\lambda^2\Big\}.
\end{aligned}\tag{42}$$

Hence, combining (41) and (42) with the help of Lemma 12, we arrive at

$$\begin{aligned}
\mathbf{E}\exp\Big\{\lambda\sqrt{\Delta_2^\alpha(\mu,X,\Xi)}\Big\}&\le \exp\Big\{\lambda\sqrt{\mathbf{E}\Delta_2^\alpha(\mu,X,\Xi)}\\
&\quad+K\lambda^2\Big[\alpha L\big[\mu,H^\alpha(X)\big]+\big\|X^\top\big[I-H^\alpha(X)\big]\mu\big\|^2+\alpha\epsilon^2\Big]\Big\}.
\end{aligned}$$

Finally, combining (36), (39), (40), and this inequality with the help of Lemma 12, we complete the proof (38). ■

2.3.1 Proof of the theorem

We make use of the following equation (see (26)):

$$\begin{aligned}
&\widetilde R_\sigma^\alpha(\beta,X,\zeta)=L[\mu,G^\alpha(X,\zeta)]\\
&-\frac{2\sigma^2}{\alpha}\mu^\top[I-H^\alpha(X)]^\top[I-H^\alpha(X)]\zeta(X)\zeta^\top(X)[I-H^\alpha(X)]\mu\\
&+\frac{2\sigma^2}{\alpha^2}\mu^\top[I-H^\alpha(X)]^\top[I-H^\alpha(X)]\zeta(X)[I-H^\alpha(X)]\zeta(X)^\top[I-H^\alpha(X)]\mu\\
&+\frac{2\epsilon^2\sigma^2}{\alpha}\mathrm{tr}\Big\{H_X^{\alpha\top}[I-H^\alpha(X)]\zeta(X)\zeta^\top(X)[I-H^\alpha(X)]\Big\}\\
&-\frac{2\epsilon^2\sigma^2}{\alpha^2}\mathrm{tr}\Big\{H_X^{\alpha\top}[I-H^\alpha(X)]\zeta(X)(I-H^\alpha(X))\zeta(X)^\top[I-H^\alpha(X)]\Big\}.
\end{aligned}$$

Thus, we obviously have

$$\begin{aligned}
\widetilde R_\sigma^\alpha(\beta,X,\zeta)&\le L[\mu,G^\alpha(X,\zeta)]+\frac{2\epsilon^2\sigma^2}{\alpha}\mathrm{tr}\big\{H^\alpha(X)^\top\zeta(X)\zeta^\top(X)\big\}\\
&\quad+\frac{2\sigma^2}{\alpha^2}\bar\mu^\top[I-h^\alpha(X)]^2\Xi(X)[I-h^\alpha(X)]\Xi(X)^\top[I-h^\alpha(X)]\bar\mu.
\end{aligned}\tag{43}$$

Since by (27)

$$\mathrm{tr}\big[H^\alpha(X)^\top \zeta(X)\zeta^\top(X)\big] = \sum_{k=1}^p h_k^\alpha(X)\|e_k^{*\top}\zeta(X)\|^2,$$

and $e_k^{*\top}\zeta(X),\ e_l^{*\top}\zeta(X),\ k \neq l$, are independent Gaussian vectors, we obtain by Condition 1 and Lemma 11

$$\begin{aligned}&\mathbf{E}\exp\Big\{\lambda\sqrt{\mathrm{tr}\big[H^\alpha(X)^\top\zeta(X)\zeta^\top(X)\big]}\Big\}\\ &\qquad\le \exp\Big\{\lambda\Big[pQ\sum_{k=1}^p [h_k^\alpha(X)]^2\Big]^{1/2}+\lambda^2\Big\}.\end{aligned} \tag{44}$$

So, our final step is to control the last term at the right-hand side in (43), but it was already done in (41). Therefore (see (44)) the additional terms at the right-hand side in (43) do not affect the concentration of $L[\mu, G^\alpha(X,\zeta)]$ obtained in Lemma 6. ■

3 Appendix

3.1 Ordered processes

Here we collected some auxiliary probabilistic facts used in deriving the concentration inequality (8).

Let $\xi(t,s),\ t\in[0,T],\ s\in[0,S]$, be a separable random field and $\sigma^2(t,s) : [0,T]\times[0,S]\to\mathbb{R}^+$ be a continuous function satisfying

Condition 2 *For all* $0\le t_1\le t_2\le T$ *and* $s\in[0,S]$

$$0\le \sigma^2(t_2,s)-\sigma^2(t_1,s)\le \sigma^2(t_2,S)-\sigma^2(t_1,S);$$

and for all $0\le s_1\le s_2\le S$ *and* $t\in[0,T]$

$$0\le \sigma^2(t,s_2)-\sigma^2(t,s_1)\le \sigma^2(T,s_2)-\sigma^2(T,s_1).$$

The following lemma generalizes Lemma 1 in [9].

Lemma 7 *Suppose Condition 2 holds. Then for any* $\lambda>0$

$$\begin{aligned}&\log\mathbf{E}\exp\Big\{\lambda \max_{t\in[0,T],\,s\in[0,S]}\frac{\xi(t,s)-\xi(0,0)}{\sqrt{\sigma^2(T,S)-\sigma^2(0,0)}}\Big\}\le \frac{\log(2)}{C_\circ}+\\ &+C_\circ \max_{0\le t_1\le t_2\le T,\,s\in[0,S]}\log\mathbf{E}\exp\Big\{\frac{\sqrt{2}}{C_\circ}\lambda\frac{\xi(t_2,s)-\xi(t_1,s)}{\sqrt{\sigma^2(t_2,s)-\sigma^2(t_1,s)}}\Big\}\\ &+C_\circ \max_{0\le s_1\le s_2\le S,\,t\in[0,T]}\log\mathbf{E}\exp\Big\{\frac{\sqrt{2}}{C_\circ}\lambda\frac{\xi(t,s_2)-\xi(t,s_1)}{\sqrt{\sigma^2(t,s_2)-\sigma^2(t,s_1)}}\Big\},\end{aligned} \tag{45}$$

where $C_\circ=(\sqrt{2}-1)/2$.

Proof. First, we construct on $[0,T]$ a family of dyadic set G_T^k, $k = 0,1,\ldots$. The set G_T^0 consist of the single point $t_0^0 = 0$. The set G_T^1 contains two points: $t_0^1 = t_0^0$ and t_1^1 defined by

$$\sigma^2(t_1^1, S) - \sigma^2(t_1^0, S) = \sigma^2(T, S) - \sigma^2(t_1^1, S),$$

or, equivalently,

$$\sigma^2(t_1^1, S) = \frac{\sigma^2(T,S) + \sigma^2(0,0)}{2}.$$

Next, if the set $G_T^k = \{t_0^k, t_1^k, \ldots, t_{2^k-1}^k\}$ has been constructed, then $G_T^{k+1} = \{t_0^{k+1}, t_1^{k+1}, \ldots, t_{2^{k+1}-1}^{k+1}\}$ is defined as follows:

$$t_{2l}^{k+1} = t_l^k, \ l = 0,1,\ldots,$$

and the points t_{2l+1}^{k+1} are computed as

$$\sigma^2(t_{2l+1}^{k+1}, S) = \frac{\sigma^2(t_l^k, S) + \sigma^2(t_{l-1}^k, S)}{2}.$$

Doing similarly, we construct also the family of the dyadic sets G_S^k, $k = 0,1,\ldots$ in $[0,S]$.

The set $G^k \in [0,T] \times [0,S]$ is defined by

$$G^k = G_T^k \times G_S^k.$$

Integer k in this definition is called set level.

For a given set level k define the following function $g^k(t,s) : [0,T] \times [0,S] \to G^k$ by

$$g^k(t,s) = \{g_1^k(t,s), g_2^k(t,s)\} = \arg \min_{\{u,v\} \in G^k : u \le t, \, v \le s} \big\{(t-u) + (s-v)\big\}.$$

In other words, this function relates (t,s) from $[0,T] \times [0,S]$ and the nearest point from G^k with coordinates less than t and s respectively.

Let us fix an integer N and let $\{t^N, s^N\}$ be an arbitrary point in G^N. Then descending from the set G^N to G^{N-1}, we obtain

$$\begin{aligned}\xi(t^N, s^N) - \xi(0,0) =& \xi(t^N, s^N) - \xi[g_1^{N-1}(t^N, s^N), g_2^{N-1}(t^N, s^N)] \\ &+ \xi[g_1^{N-1}(t^N, s^N), g_2^{N-1}(t^N, s^N)] - \xi(0,0)\end{aligned}$$

and continue the descend, we find a sequence of points such that $\{t^k, s^k\} \in G^k$ such that

$$\begin{aligned}\xi(t^N, s^N) - \xi(0,0) =& \xi(t^N, s^N) - \xi(t^{N-1}, s^{N-1}) \\ &+ \ldots + \xi(t^k, s^k) - \xi(t^{k-1}, s^{k-1}) \\ &+ \ldots + \xi(t^1, s^1) - \xi(0,0).\end{aligned} \tag{46}$$

Notice that since $t^{k-1} \le t^k, \quad s^{k-1} \le s^k$, we have

$$\begin{aligned}
\xi(t^k, s^k) - \xi(t^{k-1}, s^{k-1}) &= \frac{\xi(t^k, s^k) - \xi(t^k, s^{k-1})}{\sqrt{\sigma^2(t^k, s^k) - \sigma^2(t^k, s^{k-1})}} \\
&\times \sqrt{\sigma^2(t^k, s^k) - \sigma^2(t^k, s^{k-1})} \\
&+ \frac{\xi(t^k, s^{k-1}) - \xi(t^{k-1}, s^{k-1})}{\sqrt{\sigma^2(t^k, s^{k-1}) - \sigma^2(t^{k-1}, s^{k-1})}} \times \sqrt{\sigma^2(t^k, s^{k-1}) - \sigma^2(t^{k-1}, s^{k-1})}.
\end{aligned} \tag{47}$$

From the construction of G^k and Condition 2 we obtain the following equations:

$$\begin{aligned}
\sigma^2(t^k, s^k) - \sigma^2(t^k, s^{k-1}) &= 2^{-k+1}[\sigma^2(T, S) - \sigma^2(0, 0)], \\
\sigma^2(t^k, s^{k-1}) - \sigma^2(t^{k-1}, s^{k-1}) &= 2^{-k+1}[\sigma^2(T, S) - \sigma^2(0, 0)].
\end{aligned} \tag{48}$$

Next, combining Equations (45)-(48), we arrive at

$$\begin{aligned}
&\frac{\xi(t^N, s^N) - \xi(0, 0)}{\sqrt{\sigma^2(T, S) - \sigma^2(0, 0)}} \\
&\quad \le \sqrt{2} \sum_{k=1}^{N} 2^{-k/2} \max_{(t,s) \in G^k} \frac{\xi(t, s) - \xi[t, g_2^{k-1}(t, s)]}{\sqrt{\sigma^2(t, s) - \sigma^2[t, g_2^{k-1}(t, s)]}} \\
&\quad \le \sqrt{2} \sum_{k=1}^{N} 2^{-k/2} \max_{(t,s) \in G^k} \frac{\xi(t, s) - \xi[g_1^{k-1}(t, s), s)]}{\sqrt{\sigma^2(t, s) - \sigma^2[g_1^{k-1}(t, s), s]}}.
\end{aligned} \tag{49}$$

In order to derive from this equation (45), we make use of the convexity of $\exp(x)$, more precisely, the following inequality:

$$\log \mathbf{E} \exp\Bigg\{ \sum_{k=1}^{N} w(k) \zeta_k \Bigg\} \le \sum_{k=1}^{N} w(k) \log \mathbf{E} \exp(\zeta_k), \tag{50}$$

that holds true for any random variables ζ_k and any non-negative weights $w(k)$ such that $\sum_{k=1}^{N} w(k) = 1$. This equation follows immediately from

$$\mathbf{E} \exp\Bigg\{ \sum_{k=1}^{N} w(k) [\zeta_k - \log \mathbf{E} \exp(\zeta_k)] \Bigg\} \le \sum_{k=1}^{N} w(k) \mathbf{E} \exp[\zeta_k - \log \mathbf{E} \exp(\zeta_k)] = 1.$$

Applying twice (50) to (49) and using that the cardinality of G^k is 4^k, we get

$$
\begin{aligned}
&2\log \mathbf{E}\exp\left\{\lambda \frac{\xi(t^N,s^N)-\xi(0,0)}{\sqrt{\sigma^2(T,S)-\sigma^2(0,0)}}\right\} \\
&\le \log \mathbf{E}\exp\left\{2\sqrt{2}\lambda \sum_{k=1}^{N} 2^{-k/2} \max_{(t,s)\in G^k} \frac{\xi(t,s)-\xi[t,g_2^{k-1}(t,s)]}{\sqrt{\sigma^2(t,s)-\sigma^2[t,g_2^{k-1}(t,s)]}}\right\} \\
&+ \log \mathbf{E}\exp\left\{2\sqrt{2}\lambda \sum_{k=1}^{N} 2^{-k/2} \max_{(t,s)\in G^k} \frac{\xi(t,s)-\xi[g_1^{k-1}(t,s),s]}{\sqrt{\sigma^2(t,s)-\sigma^2[g_1^{k-1}(t,s),s]}}\right\} \qquad (51) \\
&\le \frac{1}{S_N}\sum_{k=1}^{N} 2^{-k/2}\log\left\{4^k \max_{t,s\ge s'} \mathbf{E}\exp\left[2\sqrt{2}S_N\lambda \frac{\xi(t,s)-\xi(t,s')}{\sqrt{\sigma^2(t,s)-\sigma^2(t,s')}}\right]\right\} \\
&+ \frac{1}{S_N}\sum_{k=1}^{N} 2^{-k/2}\log\left\{4^k \max_{t\ge t',s} \mathbf{E}\exp\left[2\sqrt{2}S_N\lambda \frac{\xi(t,s)-\xi(t',s)}{\sqrt{\sigma^2(t,s)-\sigma^2(t',s)}}\right]\right\},
\end{aligned}
$$

where

$$S_N=\sum_{k=1}^{N} 2^{-k/2} = (1+o(1))\frac{1}{\sqrt{2}-1}, \quad N\to\infty.$$

In view of separability of $\xi(t,s)$, taking the limit as $N\to\infty$ in Equation (51), we finish the proof of (45). ■

Lemma 8 *Suppose* $\zeta(t),\ t\ge 0$ *a separable random process and* $\sigma^2(t),\ t\ge 0$ *be a nondecreasing function such that* $\lim_{t\to\infty}\sigma^2(t)=\infty$ *and for all* $\lambda\in[0,\Lambda]$

$$\mathbf{E}\exp\bigl\{\lambda \max_{t\in[0,T]}\zeta(t)\bigr\} \le K\exp\bigl[K_\circ\sigma^2(T)\lambda^2\bigr], \qquad (52)$$

where $K,\ K_\circ$ *are some constants independent of* T*. Then for all* $x>0$ *and* $z\in(0,\Lambda)$

$$\mathbf{P}\left\{z\sup_{t\ge 0}\bigl[\zeta(t)-z\sigma^2(t)\bigr]\ge x\right\} \le KC(q,K_\circ)\exp\left[-\frac{x}{K_\circ(1+q)}\right], \qquad (53)$$

where $q>0$ *and* $C(q,K_\circ)$ *is a function of* q *and* $K_\circ$*, bounded from above for all* $q>0$*.*

Proof. Let $\{t_k,\ k=0,\ldots\}$ be a monotone positive sequence such that $\lim_{k\to\infty}t_k=\infty$.

With the help of the exponential Chebychev inequality and (52) we get for any $\lambda \in (0, \Lambda)$

$$\begin{aligned}
\mathbf{P}\Big\{ z \max_{t \geq 0} \big[\zeta(t) - z\sigma^2(t) \big] \geq x \Big\} &\leq \sum_{k=1}^{\infty} \mathbf{P}\Big\{ \max_{t \in [t_{k-1}, t_k]} \big[\zeta(t) - z\sigma^2(t) \big] \geq x \Big\} \\
&\leq \sum_{k=1}^{\infty} \mathbf{P}\Big\{ z \max_{t \in [0, t_k]} \big\{ \zeta(t) - z\sigma^2(t_{k-1}) \big\} \geq x \Big\} \qquad (54) \\
&\leq K \sum_{k=1}^{\infty} \exp\Big[-\frac{\lambda x}{z} - \lambda z \sigma^2(t_{k-1}) + K_\circ \lambda^2 \sigma^2(t_k) \Big].
\end{aligned}$$

Therefore choosing $\lambda = z/[K_\circ(1+q)]$ with $q > 0$ we obtain by (54)

$$\begin{aligned}
\mathbf{P}\Big\{ z \max_{t \geq 0} \big[\zeta(t) - z\sigma^2(t) \big] \geq x \Big\} &\leq K \exp\Big[-\frac{x}{K_\circ(1+q)} + \frac{z^2\sigma^2(t_1)}{K_\circ(1+q)^2} \Big] \\
&+ K \exp\Big[-\frac{x}{K_\circ(1+q)} \Big] \sum_{k=2}^{\infty} \exp\Big[-\frac{z^2\sigma^2(t_{k-1})}{K_\circ(1+q)} + \frac{z^2\sigma^2(t_k)}{K_\circ(1+q)^2} \Big]. \qquad (55)
\end{aligned}$$

Let us now define points t_k as follows:

$$t_1 = \max\big\{ s : \sigma^2(s) \leq z^{-1} \big\}, \quad t_k = \max\big\{ s : \sigma^2(s) \leq (1+q)p\sigma^2(t_{k-1}) \big\},$$

where $(1+q)^{-1} < p < 1$. Then it is clear that

$$-\frac{z^2\sigma^2(t_{k-1})}{1+q} + \frac{z^2\sigma^2(t_k)}{(1+q)^2} \leq -\frac{(1-p)z^2\sigma^2(t_{k-1})}{1+q} \leq -\frac{(1-p)[(1+q)p]^{k-1}}{1+q}.$$

Therefore (53) follows from (55) and the above inequality. ■

We say that $\mathcal{H}$ is a set of ordered sequences $h_k \in [0,1]$, $k = 1, \dots$ if for any $h, g \in \mathcal{H}$ either

$$h_k \leq g_k, \text{ for all } k = 1, 2, \dots;$$

or

$$h_k \geq g_k, \text{ for all } k = 1, 2, \dots.$$

Lemma 9 *Let ξ_k, $k = 1, \dots$ be i.i.d. standard Gaussian random variables and $\mathcal{H}$, $\mathcal{G}$ be sets of ordered sequences. Then for any given $z > 0$ and all $x \geq 0$*

$$\mathbf{P}\Big\{ \max_{g \in \mathcal{G}} \max_{h \in \mathcal{H}} \Big[\sum_{i=1}^{p} g_i h_i \xi_i \mu_i - z\|h\mu\|^2 \Big] \geq x \Big\} \leq K \exp\big(-K_\circ z x \big), \qquad (56)$$

where $K_\circ$, K are some generic constants.

Proof. Consider the following Gaussian random field:

$$\xi(h,g) = \sum_{i=1}^{p} g_i h_i \xi_i \mu_i, \quad h \in \mathcal{H},\, g \in \mathcal{G}.$$

Since $\xi(h,g)$ is a Gaussian random variable, we have for any $\lambda \in \mathbb{R}^+$

$$\begin{aligned}
\mathbf{E}\exp\bigl\{\lambda[\xi(h_1,g)-\xi(h_2,g)]\bigr\} &= \exp\biggl[\frac{\lambda^2\sigma^2(h_1-h_2,g)}{2}\biggr],\\
\mathbf{E}\exp\bigl\{\lambda[\xi(h,g_2)-\xi(h,g_1)]\bigr\} &= \exp\biggl[\frac{\lambda^2\sigma^2(h,g_2-g_1)}{2}\biggr],
\end{aligned} \tag{57}$$

where

$$\sigma^2(h,g) = \sum_{k=1}^{p} h_k^2 g_k^2 \mu_k^2.$$

Notice that for any $h_1,\, h_2 \in \mathcal{H}$ such that $h_2 \geq h_1$

$$\begin{aligned}
\sigma^2(h_2,g) - \sigma^2(h_1,g) &= \sum_{k=1}^{p}[h_{2,k}^2 - h_{1,k}^2] g_k^2 \mu_k^2\\
&= \sum_{k=1}^{p}[h_{2,k} - h_{1,k}][h_{2,k} + h_{1,k}] g_k^2 \mu_k^2\\
&\geq \sum_{k=1}^{p}[h_{2,k} - h_{1,k}][h_{2,k} - h_{1,k}] g_k^2 \mu_k^2 = \sigma^2(h_2-h_1,g)
\end{aligned} \tag{58}$$

and similarly for any $g_1,\, g_2 \in \mathcal{G}$ such that $g_2 \geq g_1$

$$\sigma^2(h,g_2) - \sigma^2(h,g_1) \geq \sigma^2(h,g_2-g_1). \tag{59}$$

Therefore we can apply Lemma 7 because Condition 2 for $\sigma^2(h,g)$ is obviously fulfilled. So, by this lemma and Equations (57)-(59) we obtain the following inequality:

$$\mathbf{E}\exp\biggl\{\lambda \max_{h' \geq h;\, g \in \mathcal{G}} \frac{\xi(h',g)}{\sqrt{\sigma^2(h,1)}}\biggr\} \leq K\exp\bigl(K_\circ \lambda^2\bigr),$$

which holds true for any $\lambda \in \mathbb{R}^+$. Here $K_\circ$ and K are some constants.

This equation is equivalent to

$$\mathbf{E}\exp\bigl\{\lambda \max_{h' \leq h;\, g} \xi(h',g)\bigr\} \leq K\exp\bigl[K_\circ \lambda^2 \sigma^2(h,1)\bigr], \quad \lambda \geq 0. \tag{60}$$

and thus (56) follows from (60) and Lemma 8. ■

Lemma 10 *Suppose $\mathcal{H}$ is a set of ordered sequences and ξ_k, $k = 1, \dots$ are i.i.d. standard random variables. Then for any $\gamma > 0$ and any $x \geq 0$*

$$\mathbf{P}\biggl\{\max_{h,g\in\mathcal{H}}\biggl[\sum_{k=1}^{n} g_k h_k(1-\xi_k^2) - \gamma\|h\|^2\biggr] \geq x\biggr\} \leq K\exp\bigl(-K_\circ\gamma x\bigr) \tag{61}$$

and for any $\gamma \in (0, 3/16)$ and any $x \geq 0$

$$\mathbf{P}\biggl\{\max_{h,g\in\mathcal{H}}\biggl[\sum_{k=1}^{n} g_k h_k(\xi_k^2 - 1) - \gamma\|h\|^2\biggr] \geq x\biggr\} \leq K\exp\bigl(-K_\circ\gamma x\bigr); \tag{62}$$

where $K_\circ$, K are generic constants.

Proof. The proof of (61) is similar to the one of (56). Notice that for the following random field

$$\xi_-(h,g) = \sum_{k=1}^{n} g_k h_k(1-\xi_k^2)$$

and for the following function

$$\sigma^2(h,g) = \sum_{k=1}^{n} g_k^2 h_k^2.$$

the equations

$$\begin{aligned}\mathbf{E}\exp\biggl[\lambda\frac{\xi_-(h_2,g)-\xi_-(h_1,g)}{\sqrt{\sigma^2(h_2-h_1,g)}}\biggr] &\leq \exp(\lambda^2),\\ \mathbf{E}\exp\biggl[\lambda\frac{\xi_-(h,g_2)-\xi_-(h,g_1)}{\sqrt{\sigma^2(h,g_2-g_1)}}\biggr] &\leq \exp(\lambda^2)\end{aligned} \tag{63}$$

hold true for all $\lambda \geq 0$ and all $h_2 \geq h_1,\ g_2 \geq g_1$. In order to verify (63), we make use of $\log(1+x) \geq x - x^2/2$. So, we get

$$\begin{aligned}\log\mathbf{E}\exp\biggl[\lambda\frac{\xi_-(h_2,g)-\xi_-(h_1,g)}{\sqrt{\sigma^2(h_2-h_1,g)}}\biggr] &= \frac{\lambda}{\sqrt{\sigma^2(h_2-h_1,g)}}\sum_{k=1}^{p} g_k(h_{2k}-h_{1k})\\ -\frac{1}{2}\sum_{k=1}^{p}\log\biggl[1+2\lambda\frac{g_k(h_{2k}-h_{1k})}{\sqrt{\sigma^2(h_2-h_1,g)}}\biggr] &\leq \lambda^2.\end{aligned}$$

The proof of (62) is more delicate since for the random field

$$\xi_+(h,g) = \sum_{k=1}^{n} g_k h_k(\xi_k^2 - 1)$$

Equations (63) do not hold for all $\lambda \geq 0$.

To overcome this difficulty, notice that by the convexity of $\exp(x)$ we have for any $\mu \in (0,1)$

$$
\begin{aligned}
&\mathbf{E}\exp\Big\{\lambda \max_{h\le H,\, g}[\xi_+(h,g)]\Big\} \\
&\quad \le \mu \mathbf{E}\exp\Big\{\mu^{-1}\lambda \max_{h\le H,\, g}[\xi_+(h,g)-\xi_+(H,g^{\max})]\Big\} \\
&\quad + (1-\mu)\mathbf{E}\exp\Big\{(1-\mu)^{-1}\lambda\xi_+(H,g^{\max})\Big\},
\end{aligned}
\tag{64}
$$

where $g^{\max}$ is the maximal element in $\mathcal{G}$.

The main idea in this inequality is that for the following random field

$$
\xi(h,g) = \xi_+(h,g) - \xi_+(H,g^{\max})
$$

inequalities

$$
\begin{aligned}
\mathbf{E}\exp\left[\lambda\frac{\xi(h_1,g)-\xi(h_2,g)}{\sqrt{\sigma^2(h_1-h_2,g)}}\right] &\le \exp(\lambda^2), \\
\mathbf{E}\exp\left[\lambda\frac{\xi(h,g_1)-\xi_-(h,g_2)}{\sqrt{\sigma^2(h,g_1-g_2)}}\right] &\le \exp(\lambda^2)
\end{aligned}
$$

hold for all $\lambda \ge 0$. Therefore, we have

$$
\mathbf{E}\exp\big\{\lambda\mu^{-1}\max_{h\le H,\, g\in\mathcal{G}}\xi(h,g)\big\} \le K\exp\big[K_\circ\lambda^2\mu^{-2}\sigma_\circ^2(H)\big], \quad \lambda\ge 0, \tag{65}
$$

where

$$
\sigma_\circ^2(H) = \sum_{k=1}^p H_k^2, \quad \mathrm{K}_\circ = \frac{4\sqrt{2}}{\sqrt{2}-1} \le 14.
$$

Next, with the help of the inequality

$$
-\log(1-x) \le x + 8x^2, \quad x\in[0,3/4],
$$

we obtain

$$
\begin{aligned}
\log \mathbf{E}\exp\Big\{(1-\mu)^{-1}\lambda\xi_+(H,g^{\max})\Big\} &= -\frac{\lambda}{1-\mu}\sum_{k=1}^p H_k g_k^{\max} \\
-\frac{1}{2}\sum_{k=1}^p \log\left[1-\frac{2\lambda H_k g_k^{\max}}{1-\mu}\right] &\le \frac{16\lambda^2}{(1-\mu)^2}\sigma_\circ^2(H).
\end{aligned}
\tag{66}
$$

This inequality holds true if

$$
\frac{2\lambda}{1-\mu} \le \frac{3}{4}.
$$

Thus, choosing $\mu = 1/2$, we obtain from (64)-(66) that for all $\lambda \in [0,3/16]$

$$
\log \mathbf{E}\exp\Big[\lambda\max_{h\le H,\, g}\xi_+(h,g)\Big] \le 64\lambda^2\sigma^2(H) + K.
$$

Therefore (62) follows from Lemma 8. ■

3.2 Concentration inequalities for quadratic forms in Gaussian random variables

The main idea in the proofs of Theorems 3 and 4 is based on simple facts about distributions of quadratic forms in Gaussian random variables. They are undoubtedly known and we provide their proofs here only for reader's convenience. We refer interested readers to [16] (Section 4), where more precise results can be found.

Lemma 11 *Let*

$$\zeta_b = \sum_{k=1}^{\infty} b_k \chi_k^2(q_k),$$

where $\chi_k^2(q)$ are i.i.d. standard χ^2-random variables with q degrees of freedom, $b_k \geq 0$ and $q_k \geq 1$ are deterministic sequences. Then for any $\lambda \geq 0$

$$\mathbf{E} \exp\bigl(\lambda\sqrt{\zeta_b}\bigr) \leq \exp\bigl(\lambda\sqrt{\mathbf{E}\zeta_b} + \lambda^2 \max_k b_k\bigr). \tag{67}$$

Proof. Denote for brevity

$$B = \max_k b_k, \quad M = \sum_{k=1}^{\infty} b_k q_k = \mathbf{E}\zeta_b^2.$$

Inequality (67) may be proved with the help of the following simple inequality

$$\begin{aligned}
\mathbf{E} \exp\bigl(\lambda\sqrt{\zeta_b}\bigr) &= \min_{\mu\geq 0} \int_0^{\infty} \exp(\lambda\sqrt{x} - \mu x) \exp(\mu x) p_{\zeta_b}(x)\, dx \\
&\leq \min_{\mu\geq 0} \exp\Bigl[\max_{x\geq 0}\bigl(\sqrt{x}\lambda - \mu x\bigr)\Bigr] \mathbf{E} \exp(\mu\zeta_b) \\
&= \exp\biggl\{\min_{\mu\geq 0}\biggl[\frac{\lambda^2}{4\mu} + \log \mathbf{E}\exp(\mu\zeta_b)\biggr]\biggr\}.
\end{aligned} \tag{68}$$

Since

$$\mathbf{E}\exp(\mu\zeta_b) = \exp\biggl\{-\frac{1}{2}\sum_{k=1}^{\infty} q_k \log(1 - 2b_k\mu)\biggr\},$$

we have

$$\mathbf{E} \exp\bigl(\lambda\sqrt{\zeta_b}\bigr) \leq \min_{\mu\geq 0}\biggl\{\frac{\lambda^2}{4\mu} - \frac{1}{2}\sum_{k=1}^{\infty} q_k \log(1 - 2b_k\mu)\biggr\} \tag{69}$$

The above minimum is attained obviously at $\mu_\circ$, which is a solution to

$$\frac{\lambda^2}{4\mu_\circ^2} = \sum_{k=1}^{\infty} \frac{b_k q_k}{1 - 2\mu_\circ b_k}. \tag{70}$$

Next notice that

$$-\log(1-t) \le \frac{t}{1-t}$$

and therefore

$$-\frac{1}{2}\log(1-2\mu_\circ b_k) \le \frac{\mu_\circ b_k}{1-2\mu_\circ b_k}.$$

Thus, we have in view of (70)

$$\frac{\lambda^2}{4\mu_\circ} - \frac{1}{2}\sum_{k=1}^{\infty} q_k \log(1-2h_k\mu_\circ) \le \frac{\lambda^2}{4\mu_\circ} + \sum_{k=1}^{\infty} \frac{\mu_\circ h_k q_k}{1-2\mu_\circ h_k} = \frac{\lambda^2}{2\mu_\circ}. \tag{71}$$

On the other hand, it follows from (70) that

$$\frac{\lambda^2}{4\mu_\circ^2} \le \frac{M}{1-2B\mu_\circ}$$

or, equivalently,

$$\mu_\circ \ge \left[m\left(1+\sqrt{1+\frac{4M}{\lambda^2 B}}\right)\right]^{-1}.$$

Hence

$$\frac{\lambda^2}{2\mu_\circ} \le \frac{\lambda^2 B}{2} + \frac{\lambda^2 B}{2}\sqrt{1+\frac{4M}{\lambda^2 B}} \le \lambda^2 B + \lambda\sqrt{M}$$

and combining this inequality with Equations (71) and (69), we complete the proof. ■

We will need also the following simple fact.

Lemma 12 *Let ζ_i, $i = 1, 2$, be nonnegative random variables such that for all $\lambda \ge 0$*

$$\mathbf{E}\exp\big(\lambda\sqrt{\zeta_i}\big) \le \exp\left(\lambda\sqrt{\mathbf{E}\zeta_i} + \frac{\sigma_i^2\lambda^2}{2}\right). \tag{72}$$

Then

$$\mathbf{E}\exp\big(\lambda\sqrt{\zeta_1+\zeta_2}\big) \le \exp\left[\lambda\sqrt{\mathbf{E}(\zeta_1+\zeta_2)} + \frac{(\sigma_1+\sigma_2)^2\lambda^2}{2}\right]. \tag{73}$$

If ζ_1 and ζ_2 are independent, then

$$\mathbf{E}\exp\big(\lambda\sqrt{\zeta_1+\zeta_2}\big) \le \exp\left[\lambda\sqrt{\mathbf{E}(\zeta_1+\zeta_2)} + \frac{(\sigma_1^2+\sigma_2^2)\lambda^2}{2}\right]. \tag{74}$$

Proof. Denote

$$\Delta_i = \big[\sqrt{\zeta_i} - \sqrt{\mathbf{E}\zeta_i}\big]_+,$$

where $[x]_+ = \max(x, 0)$. It follows immediately from (72) that

$$\mathbf{E}\exp(\lambda\Delta_i) \le \exp\left(\frac{\sigma_i^2\lambda^2}{2}\right). \tag{75}$$

It is also clear that

$$\zeta_i \le (\sqrt{\mathbf{E}\zeta_i} + \Delta_i)^2.$$

and therefore

$$\begin{aligned}\zeta_1+\zeta_2 \le& \mathbf{E}(\zeta_1+\zeta_2) + 2\sqrt{\mathbf{E}(\zeta_1+\zeta_2)}\max\{\Delta_1,\Delta_2\} + \Delta_1^2+\Delta_2^2\\ \le& \Bigl[\sqrt{\mathbf{E}(\zeta_1+\zeta_2)} + \sqrt{\Delta_1^2+\Delta_2^2}\Bigr]^2.\end{aligned}$$

Hence

$$\sqrt{\zeta_1+\zeta_2} \le \sqrt{\mathbf{E}(\zeta_1+\zeta_2)} + \Delta_1+\Delta_2. \tag{76}$$

Next, by Hölder's inequality and (75),

$$\begin{aligned}\mathbf{E}\exp\bigl\{\lambda(\Delta_1+\Delta_2)\bigr\} \le& \mathbf{E}^{1/p}\exp\bigl\{p\lambda\Delta_1\bigr\}\mathbf{E}^{1/q}\exp\bigl\{q\lambda\Delta_2\bigr\}\\ \le& \exp\biggl\{\frac{\lambda^2\bigl[p\sigma_1^2+q\sigma_2^2\bigr]}{2}\biggr\},\end{aligned}$$

where $1/p+1/q=1$. Therefore, minimizing the right-hand side at this equation in p, we complete the proof of (73) in view of (76). The proof of (74) is straightforward. ■

4 Acknowledgments

The research was carried out at the IITP RAS at the expense of the Russian Foundation for Sciences (project 14-50-00150).

References

[1] A. Belloni, M. Rosenbaum, A. Tsybakov. "Linear and Conic Programming Estimators in High-Dimensional Errors-in-variables Models", arXiv:1408.0241.

[2] A. Belloni, M. Rosenbaum, A. Tsybakov. "An l_1, l_2, l_∞ - Regularization Approach to High-Dimensional Errors-in-variables Models", arXiv:1412.5216.

[3] L. Cavalier and Yu. Golubev. "Risk hull method and regularization by projections of ill-posed inverse problems." *Ann. of Stat.* **34**, No. 4. 1653–1677 (2006).

[4] L. Cavalier and N. W. Hengartner. "Adaptive estimation for inverse problems noisy operators." *Inverse Problems* **21**, 1345–1361 (2005).

[5] X. Chen and M. Reiss. "On rate optimality for ill-posed inverse problems in econometrics." *Econometric Theory* **27**, 497–521 (2011).

[6] S. Efroimovich and V. Koltchinskii. "On inverse problems with unknown operators." *IEEE Trans. Inform. Theory* **47**, 2876–2894 (2001).

[7] H. W. Engl, M. Hanke, and A. Neubauer. "Regularization of Inverse Problems." *Mathematics and its Applications, 375.* (Kluwer Academic Publishers Group. Dordrecht 1996).

[8] G. H. Golub, M. Heath, and G. Wahba. "Generalized crossvalidation as a method to choosing a good ridge parameter", Technometrics, **21**, 215-223 (1979).

[9] Yu. Golubev. "On universal oracle inequalities related to high dimensional linear models." *Ann. Statist.* **38**, No. 5, 2751-2780 (2010).

[10] Yu. Golubev and D. Ostrovski, D. "Concentration Inequalities for the Exponential Weighting Method." *Math. Methods of Statist.* , **23**, No. 1, 1-18 (2014).

[11] Yu. Golubev and T. Zimolo. "Estimation in Ill-posed Linear Models with Nuisance Design", *Math. Methods of Stat.* **24**, 1-15 (2015).

[12] M. Hoffmann and M. Reiss. "Nonlinear estimation for linear inverse problems with error in the operator." *Annals of Statist.* **36**, 1, 310–336 (2008).

[13] J. Johannes, S. van Bellegem, and A. Vanhems. "Convergence rates for ill-posed inverse problems with an unknown operator." *Economic Theory*, **27**, 522–545 (2011).

[14] A. Kneip. "Ordered linear smoothers". *Ann. Statist.* **22**, 835–866, (1994).

[15] C. Marteau. "Regularization of inverse problems with unknown operator." *Math. Methods of Statist.* **15**, 415–433 (2006).

[16] A. M. Mathai and S. P. Provost. "Quadratic forms in random variables". (Marcel Dekker, New York, 1977).

[17] A. N. Tikhonov and V. A. Arsenin. "Solution of Ill-posed Problems." *Scripta Series in Mathematics.* (V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977).